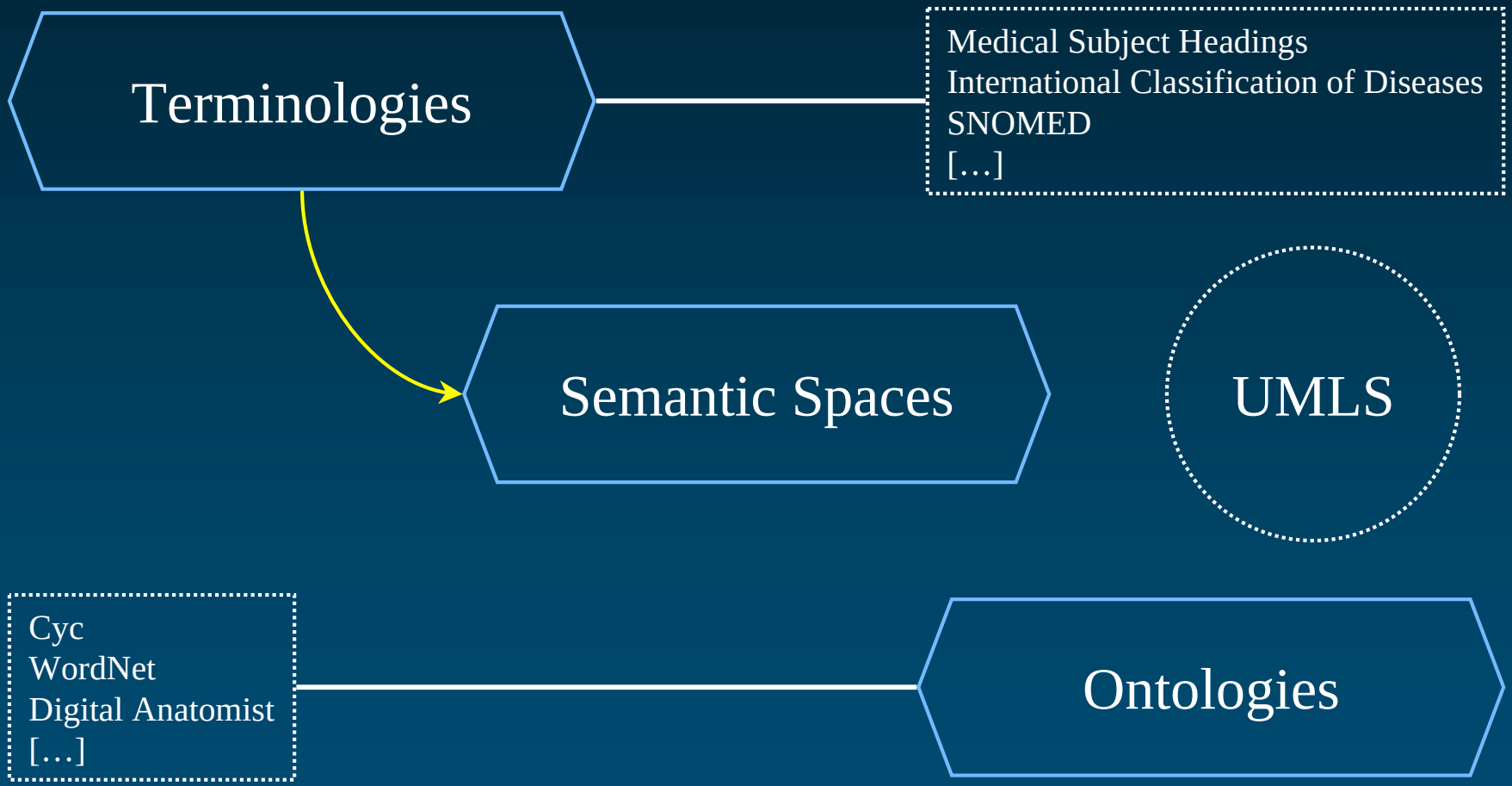
Ontological Spring
Naumburg, Germany - April 17-20, 2002

From UMLS concept spaces to a biomedical ontology

Olivier Bodenreider
National Library of Medicine
Bethesda, Maryland - USA

Anita Burgun
Medical School / Univ. Hospital
Rennes, France

Biomedical knowledge organization



Biomedical terminologies

- Core vocabularies
  - anatomy (UWDA, Neuronames)
  - drugs (First DataBank, Micromedex)
  - medical devices (UMD, SPN)
- Several perspectives
  - clinical terms (SNOMED, CTV3)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - standards (HL7, LOINC)

Biomedical terminologies (cont'd)

- Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- Knowledge bases (AI/Rheum, DXplain, QMR)

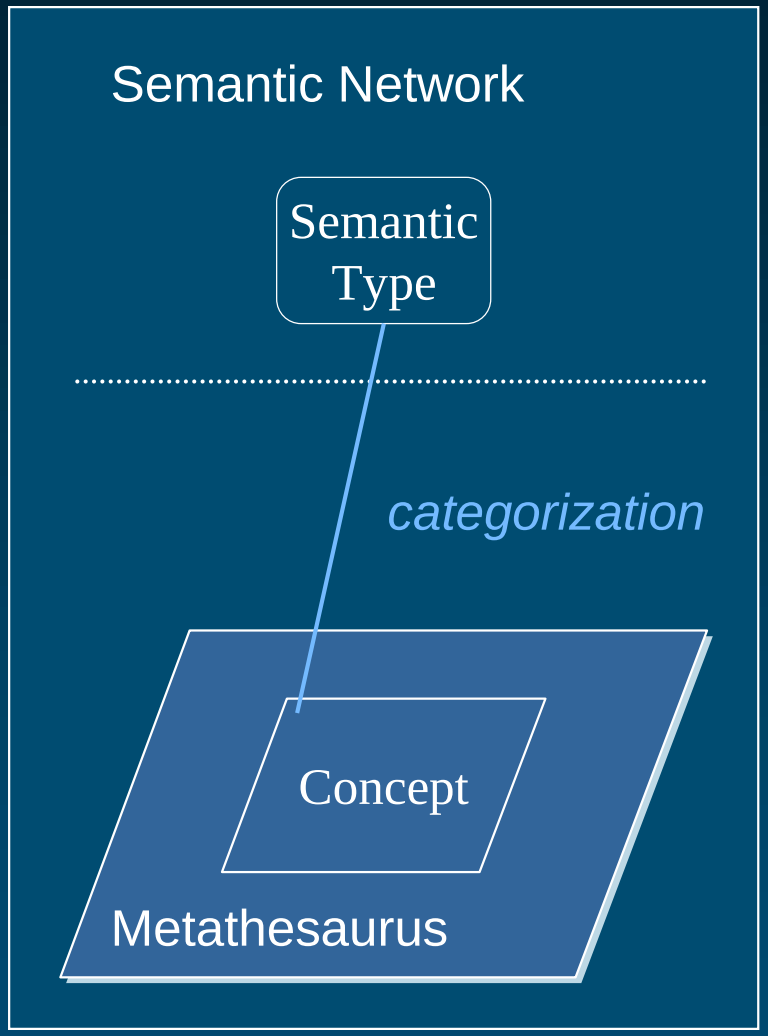
Outline

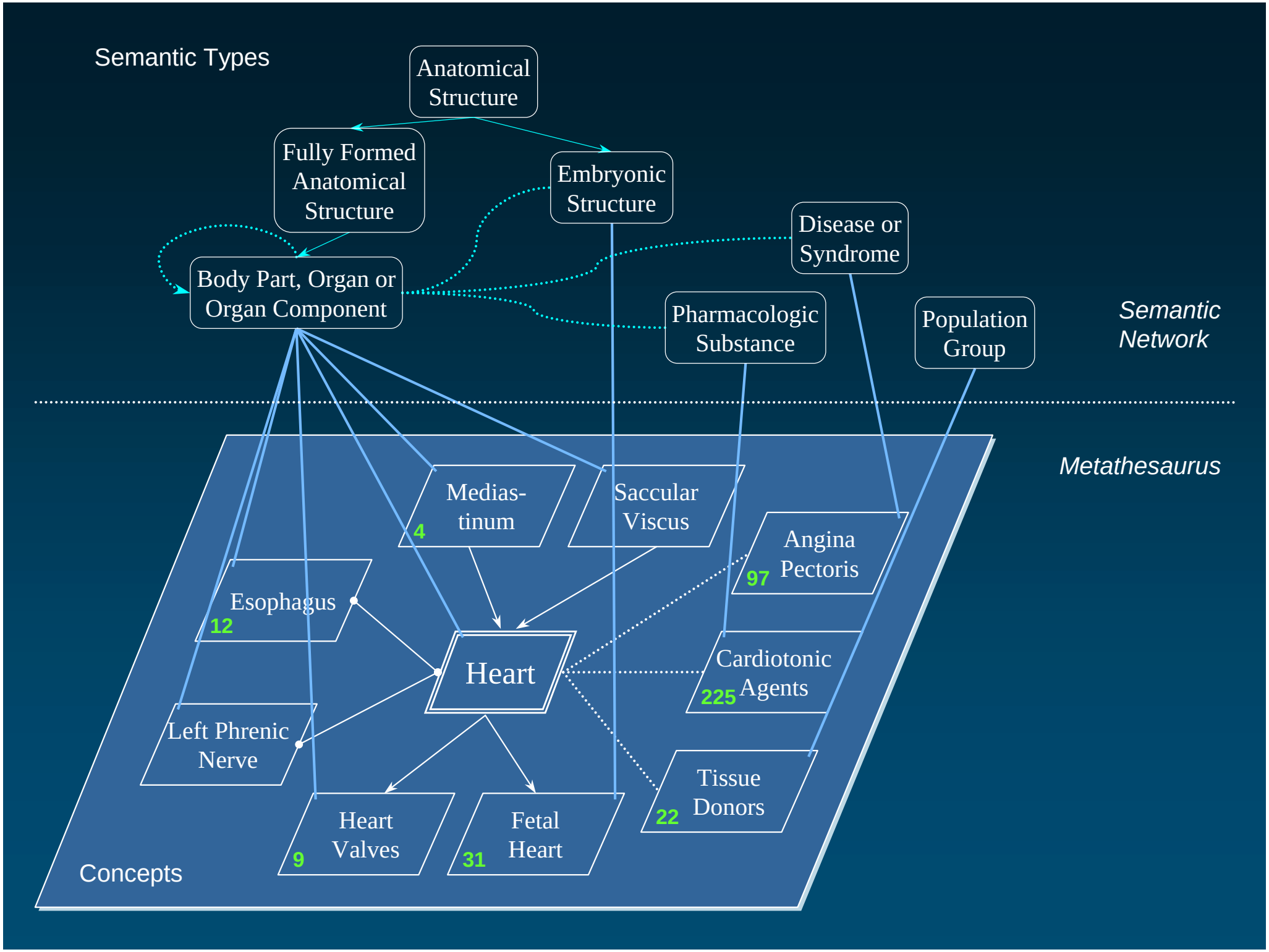
- UMLS concept spaces
- From concept spaces to ontologies
  - Structural approach
  - Semantic approach
  - Linguistic approach

UMLS concept spaces

UMLS

- Two-level structure
  - Semantic Network
    - 134 Semantic Types (STs)
    - Relationships among STs
  - Metathesaurus
    - 800,000 concepts
    - Inter-concept relationships
  - Link = categorization
    - Often isa
    - Rarely is an instance of




Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

UMLS Semantic locality

- What characterizes concepts [McCray & al., Methods 1995]
  - Textual definitions
  - Set of terms having the same meaning
  - Set of related concepts
    - Hierarchical relationships
    - Associative relationships
      - Symbolic
      - Statistical
  - Categorization (semantic types)

Semantic locality

[Nelson & al., AMIA 1992]

Concept space

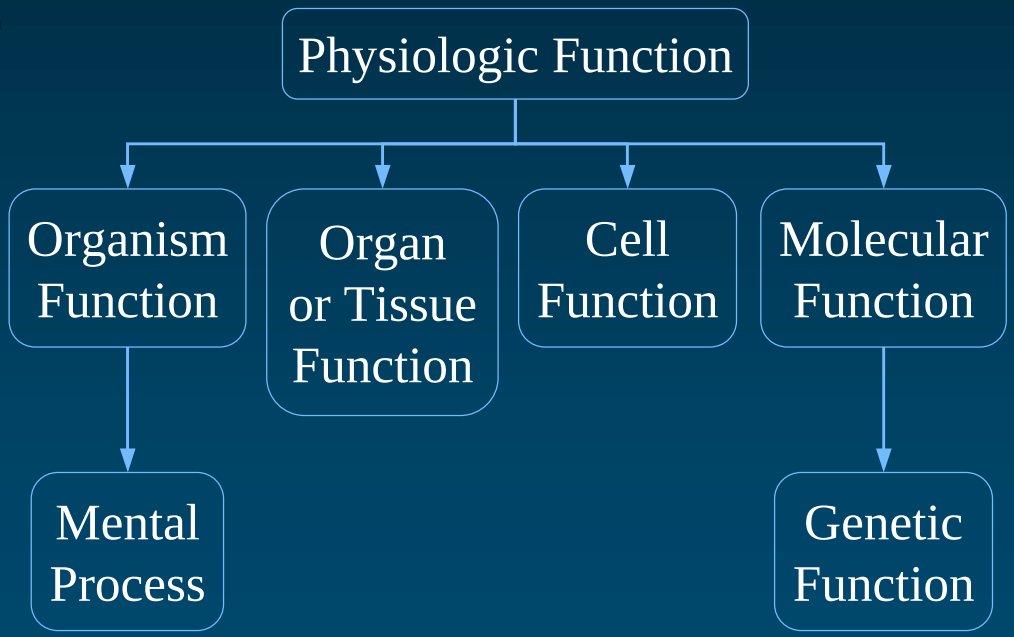
UMLS Concept spaces

- Set of related concepts
- Useful for
  - Navigation UMLS Semantic Navigator

    umlsks.nlm.nih.gov → Resources → Semantic Navigator
  - Automatic indexing/
    Information retrieval [Bodenreider & al., AMIA 1998]
- But no ontology because
  - No[t enough] definitions
  - No consistent principles used to organized concepts

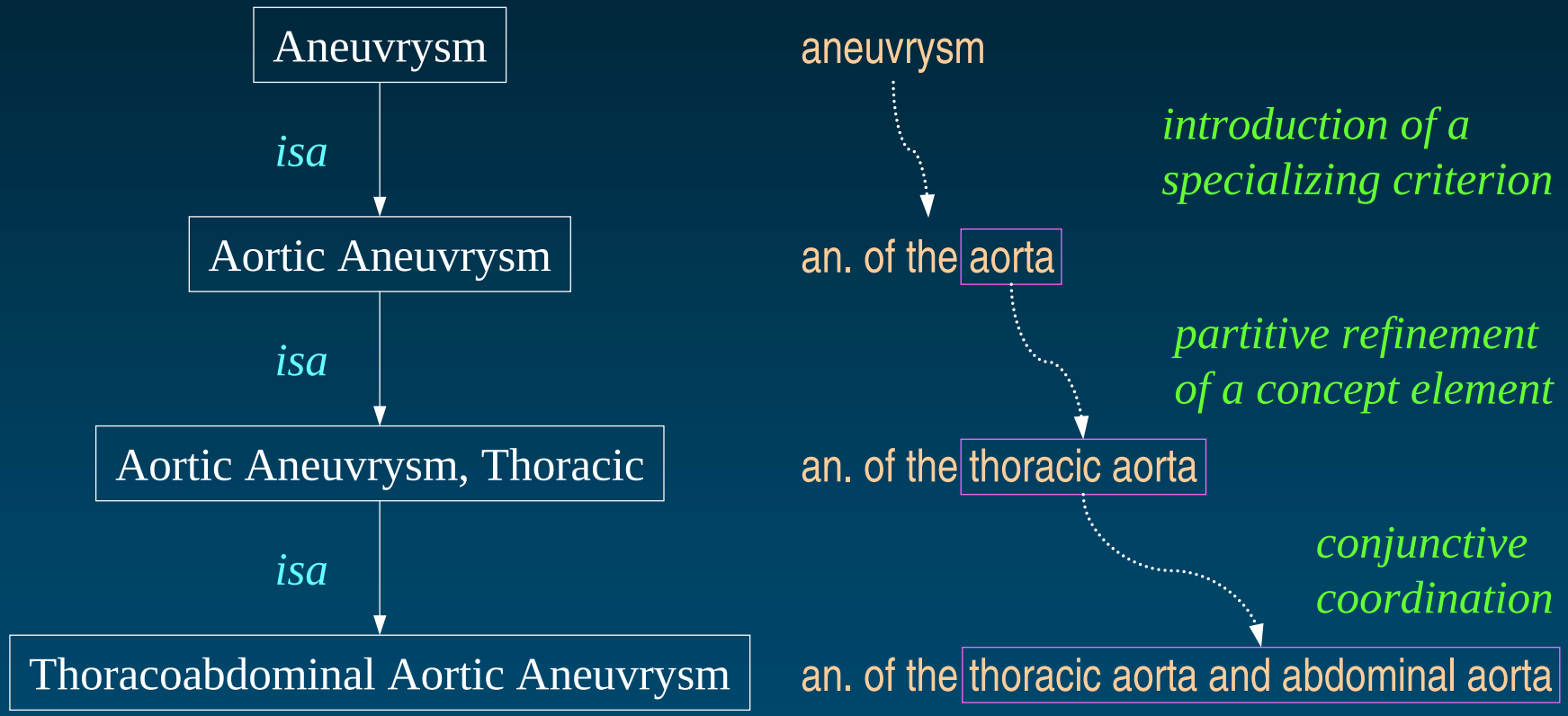
Hierarchy

- Hierarchical relationships
  - Taxonomy (isa)
  - Meronomy (part of)
- Partial ordering
  - [Reflexivity]
  - Antisymmetry
  - Transitivity
- Inheritance
- Reasoning



Principles of subsumption

[Bernauer, AMIA 1994]



Hierarchies in source vocabularies

- Structure
  - Single tree
  - Polyhierarchical (multiple parents)
- Relationships
  - Usually implicit
  - May be other than *isa* or *part of*
    - E.g., Thesaurus relationships

Hierarchies in source vocabularies

- Often task-driven
  rather than based on principles
- Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- Generally not suitable for reasoning

UMLS Biomedical ontology

- Assumption
  - The UMLS has the potential to provide [most of] the concepts and relationships needed in a biomedical ontology
- Ontology requirements
  - Relationships must be explicit
  - Relationships must be principled
  - Relationships must be consistent
- Problem
  - Select UMLS relationships compatible with ontology requirements

From concept spaces to ontologies

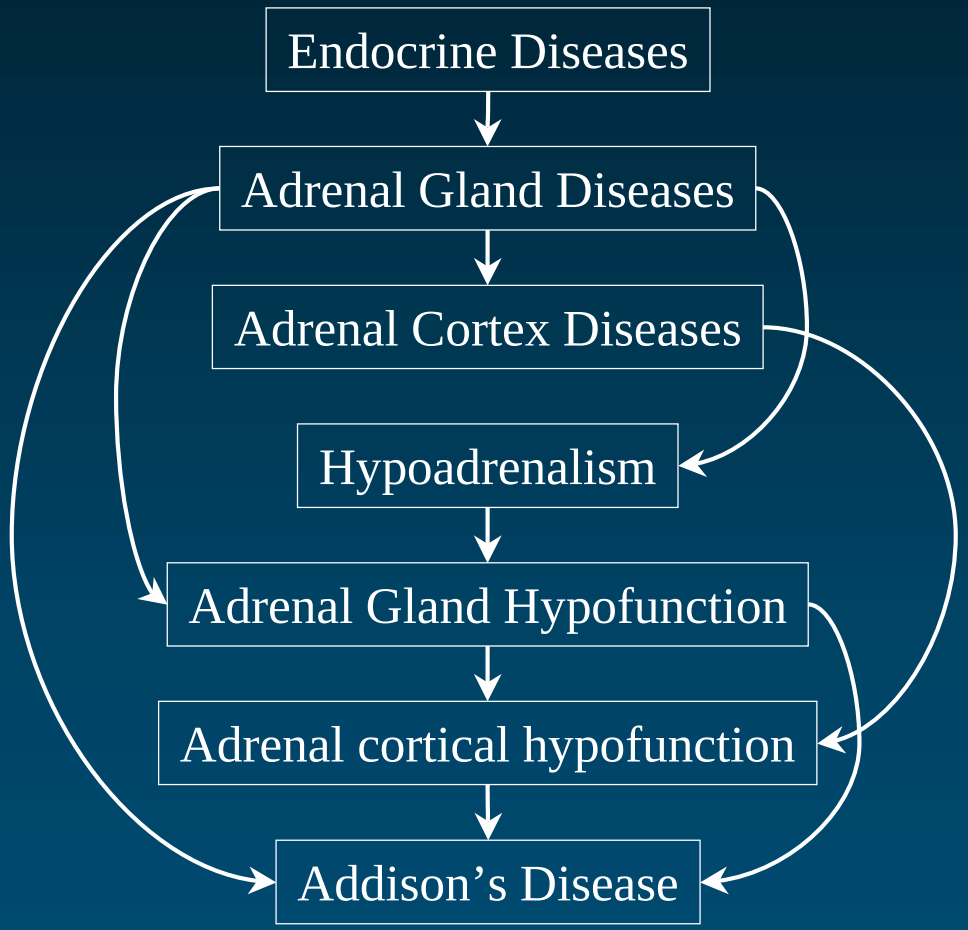
Several approaches

- Structural approach
  - Properties of partial ordering relations
- Semantic approach
  - Consistency between relationships in the Semantic Network and in the Metathesaurus
- Linguistic approach
  - Properties of adjectival modification

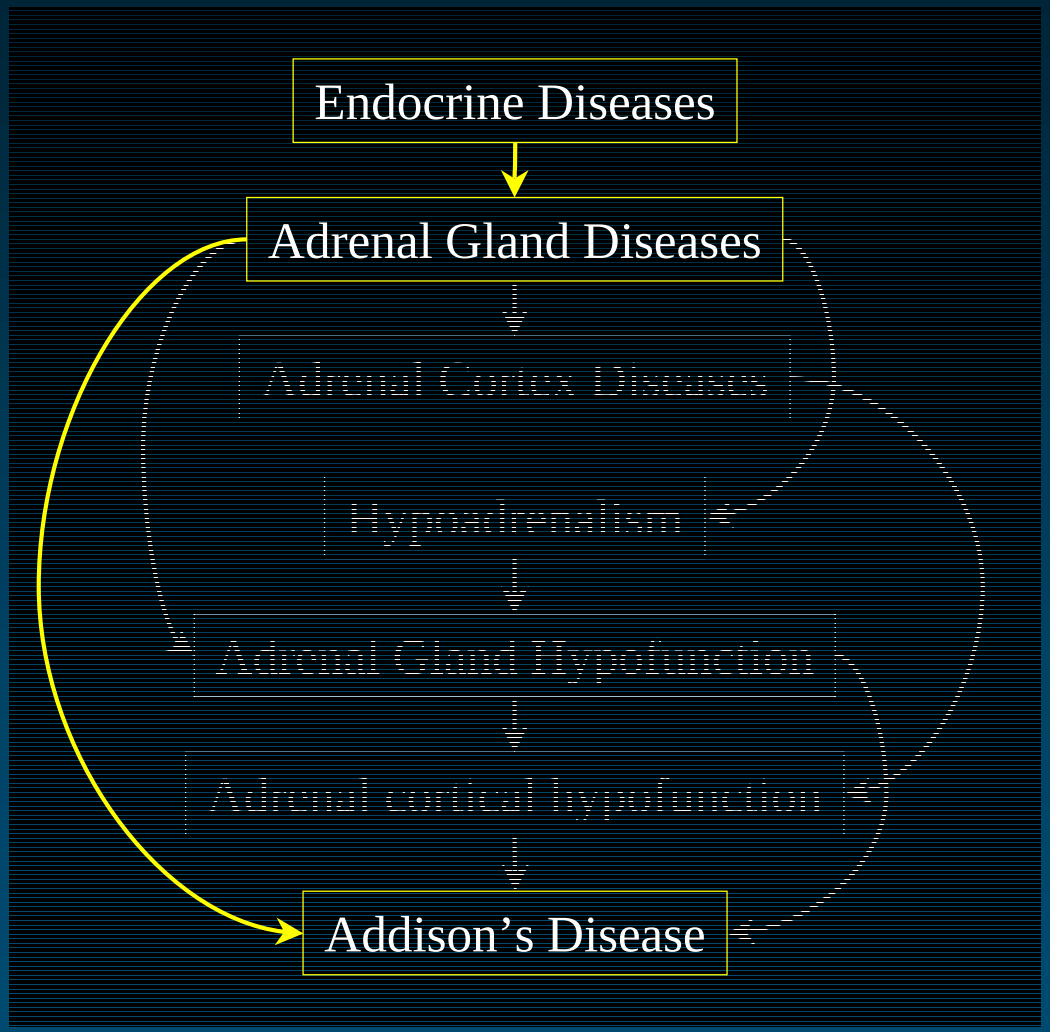
From concept spaces to ontologies

1. Structural approach

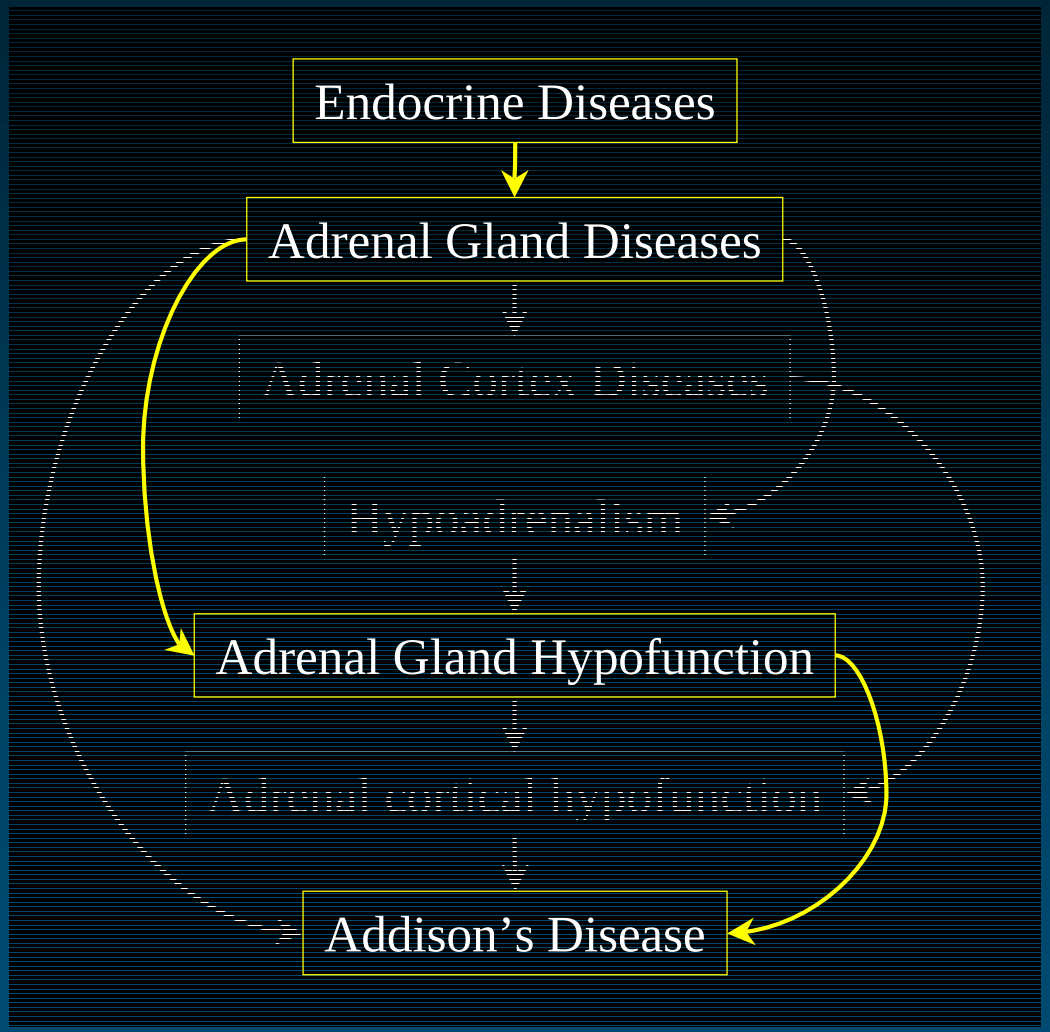
AD in UMLS Contexts



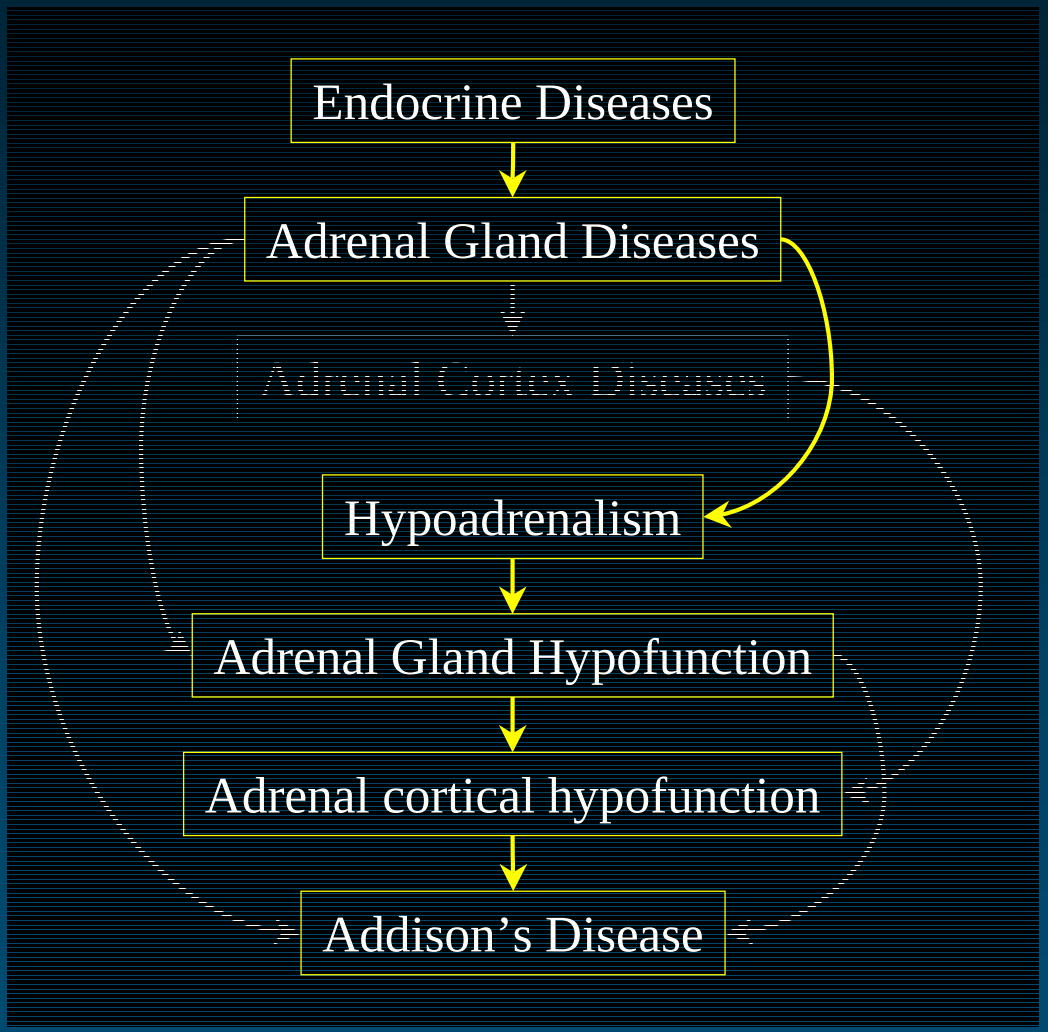
AD in UMLS SNOMED context



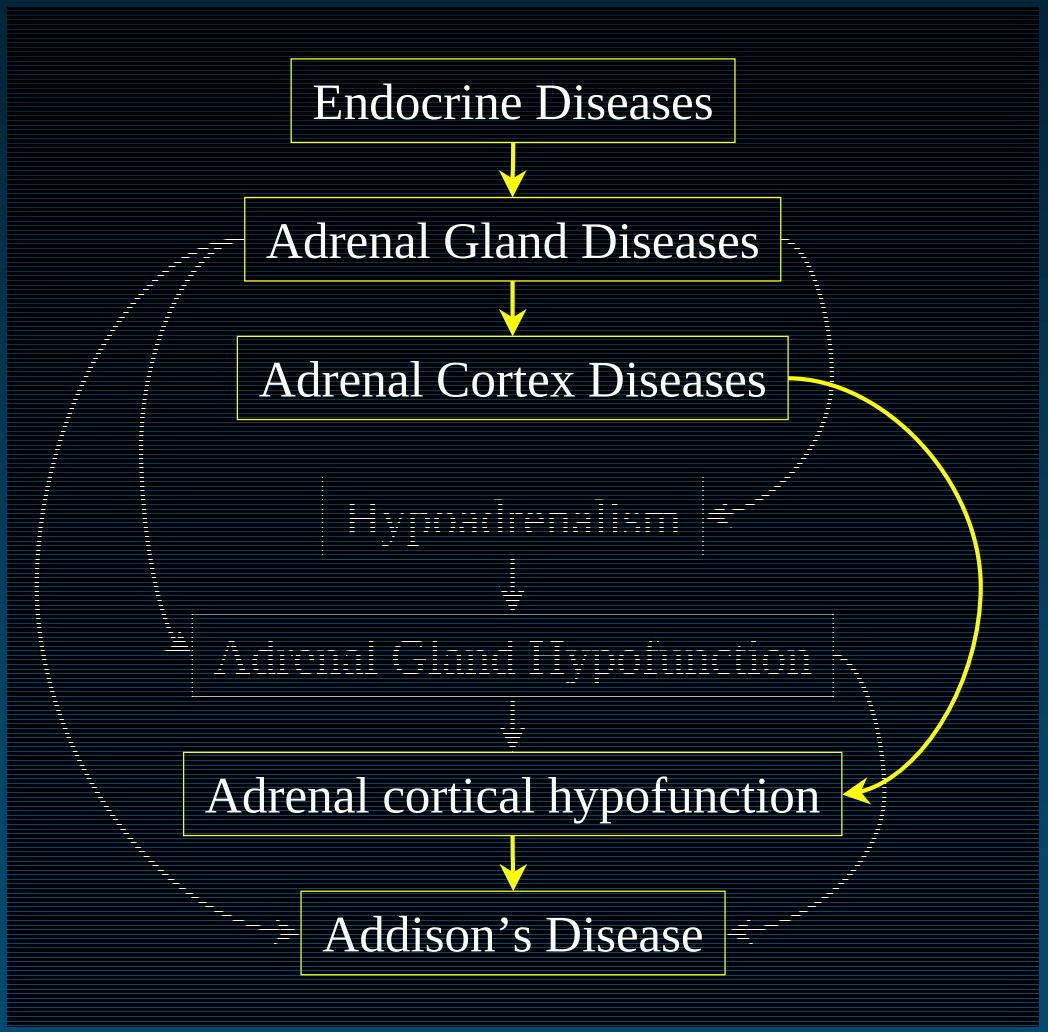
AD in UMLS MeSH context



AD in UMLS Read Codes context

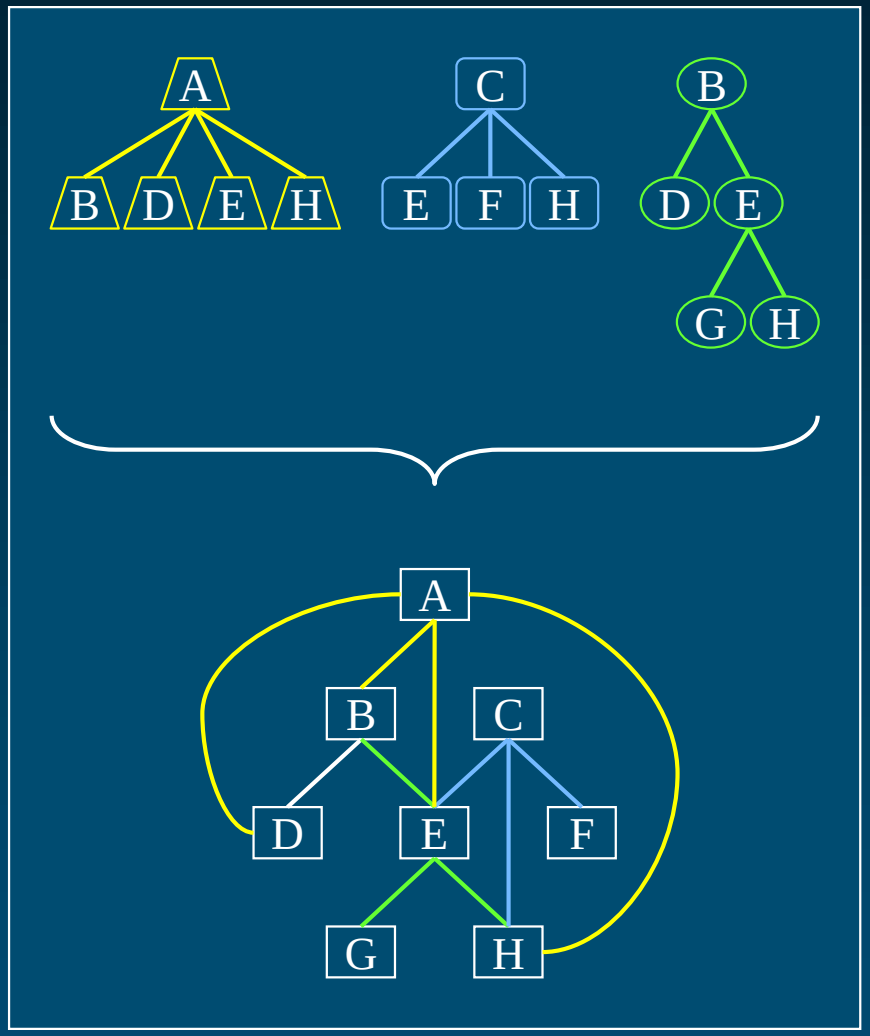


AD in UMLS AOD Thes. context

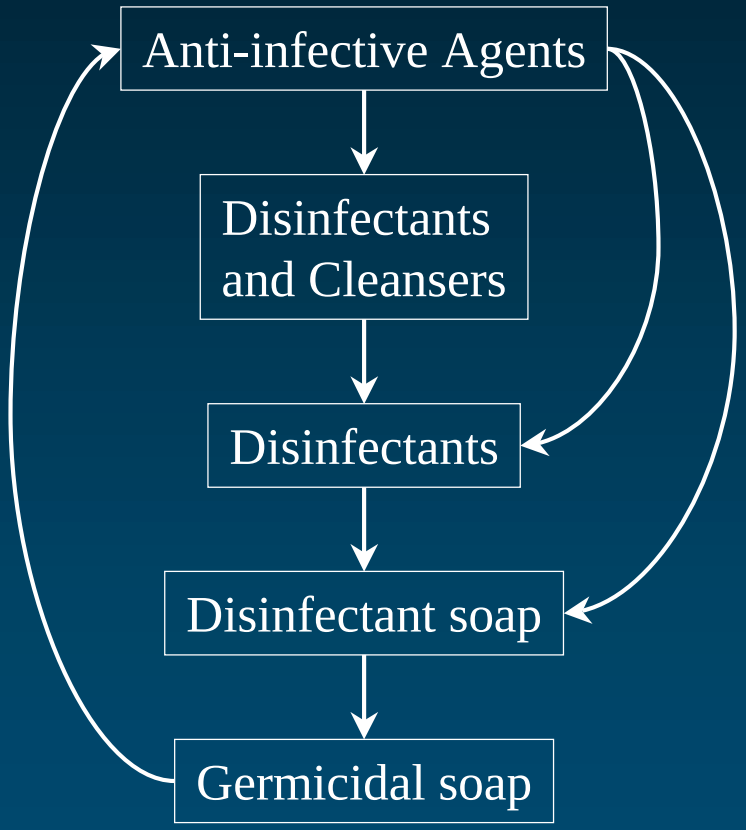


Hierarchical relationships in the UMLS

- Origin
  - Inherited from source vocabularies
    - Called Parent / Child
  - Specifically generated
    - Called Broader / Narrower
- Combined hierarchies
  - Possibly heterogeneous semantics
  - Directed acyclic graph (DAG) structure

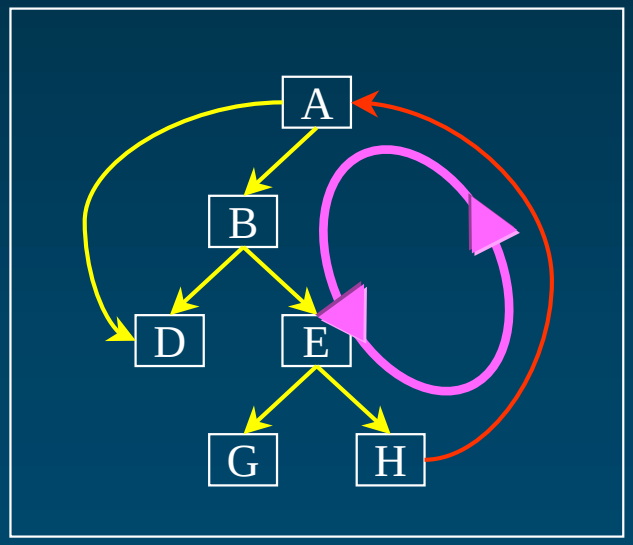


Actually, there are some cycles



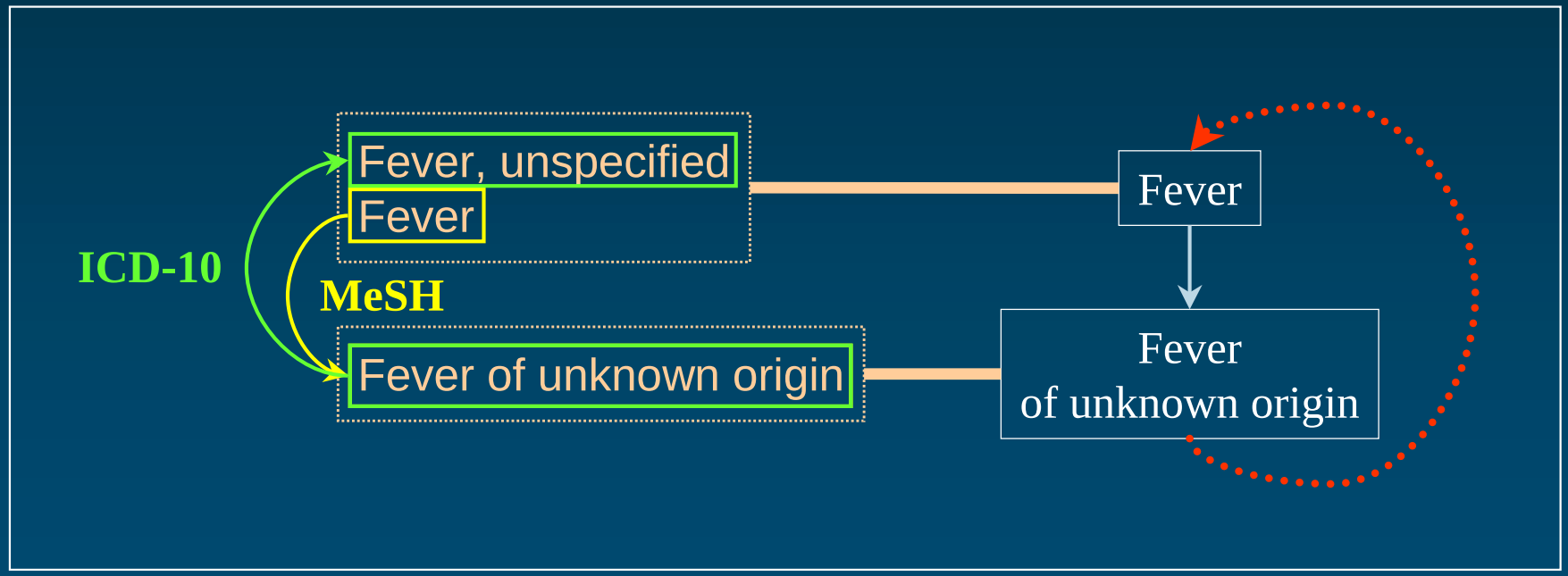
Issues with cycles

- Theoretical
  - Violate the antisymmetry property of partial ordering relations
- Practical
  - Loops in graph traversal



Cycle due to underspecification

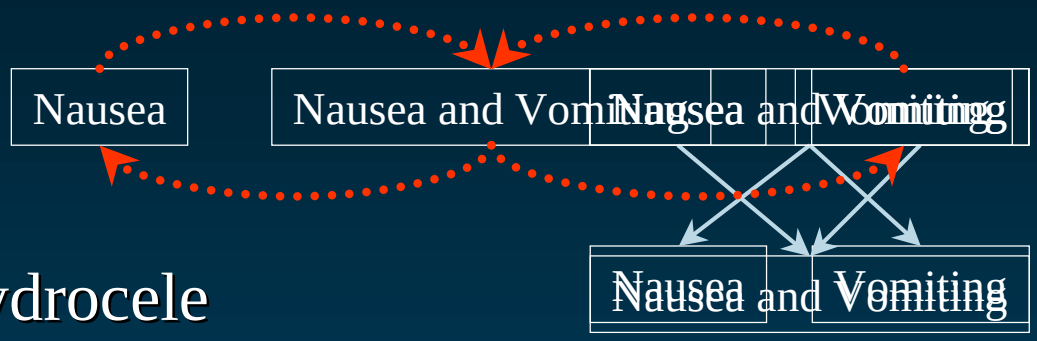
- Specified and underspecified terms
  - May appear at different levels in a source hierarchy
  - Are clustered into the same concept (same meaning)



Other causes

[Bodenreider, AMIA 2001]

- Compound terms
- Metadata
  - HYDROCELE, Hydrocele
- Classes and member
  - Purines, Purine
- Organizational conventions
  - Acid + Base ⇄ Salt + Water
- Idiopathic
  - Wrong relationships
  - Use of non-hierarchical relationships in “hierarchies”



From concept spaces to ontologies

2. Semantic approach

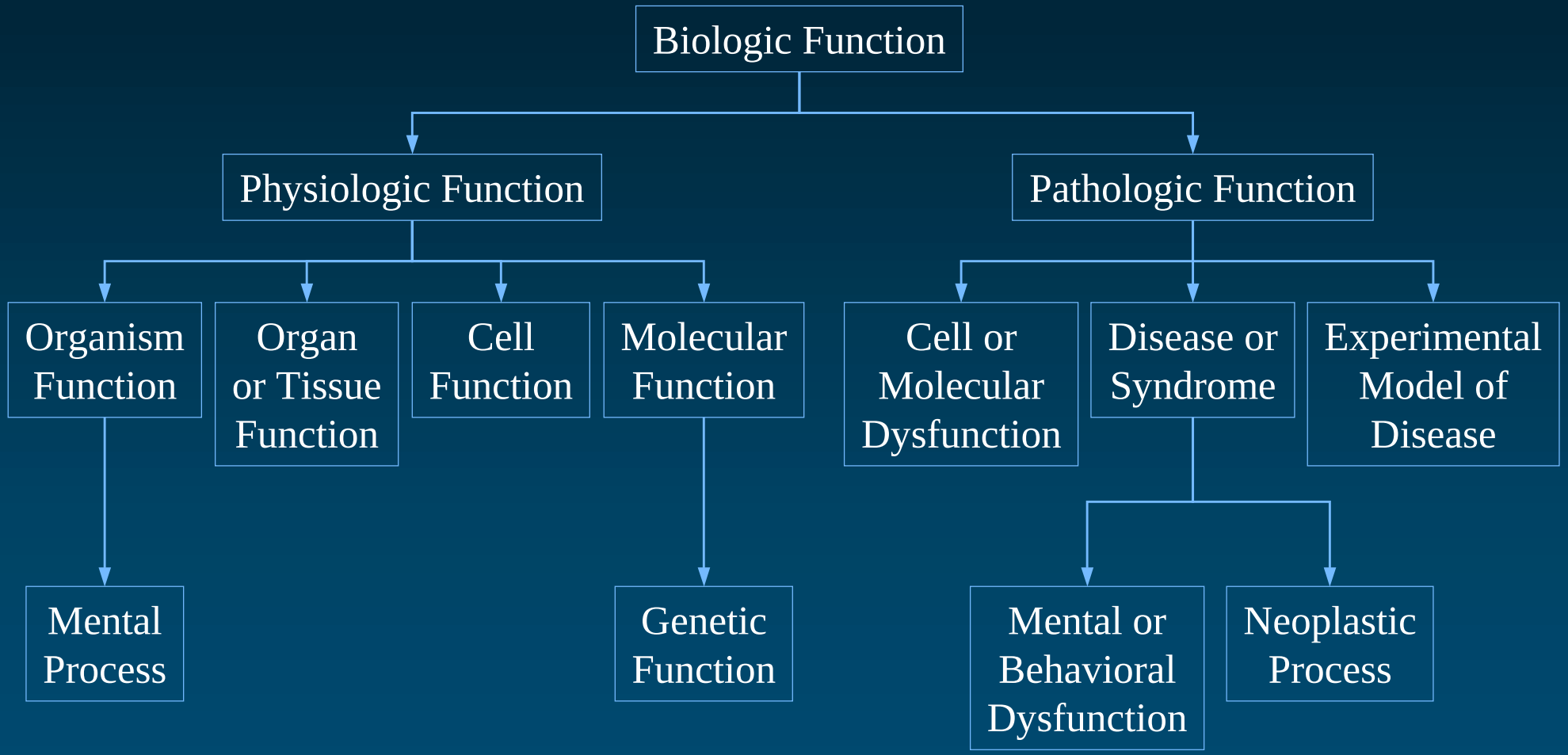
Semantic Network

- Semantic types (134)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process

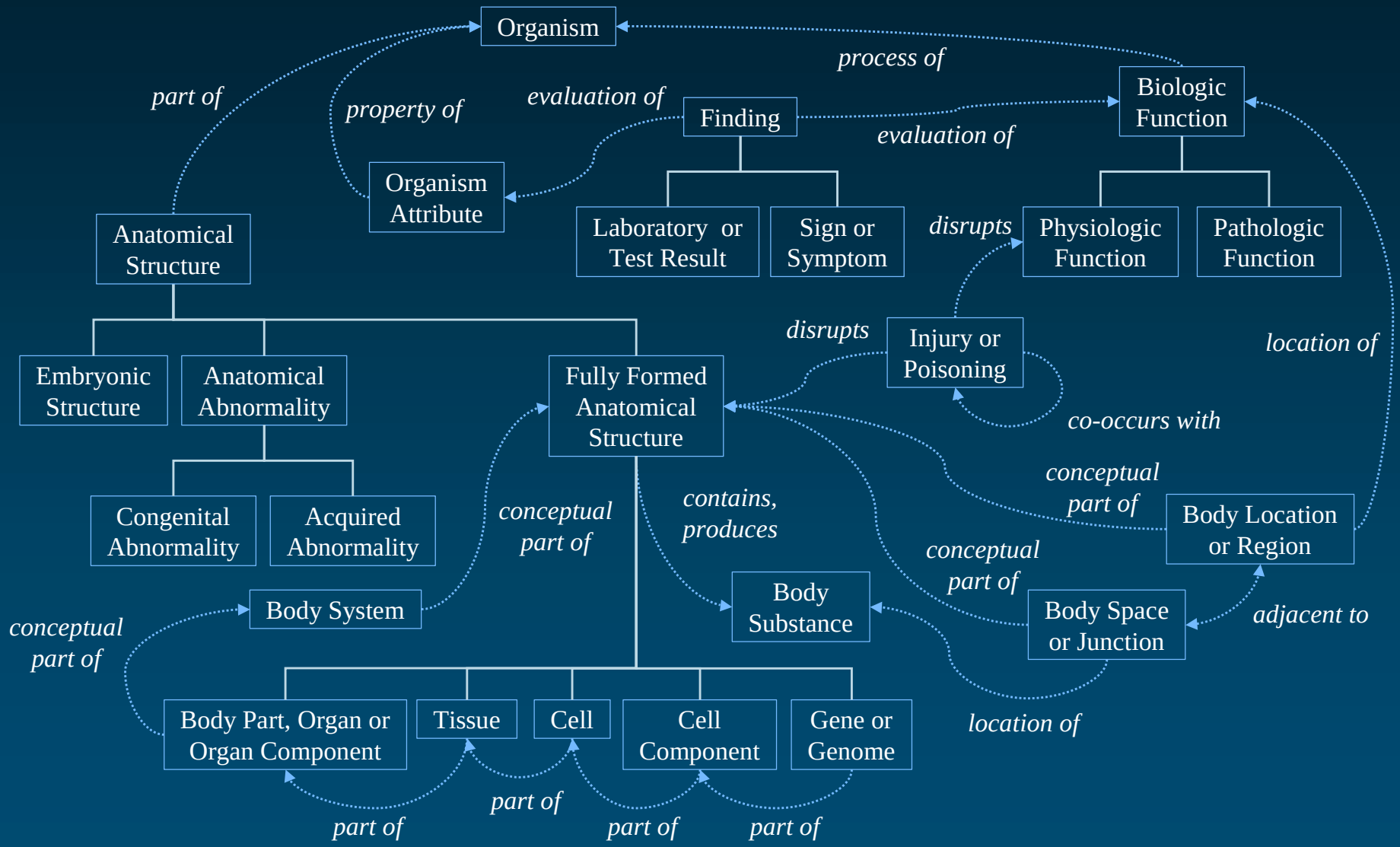
Semantic Network

- Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - Animal *isa* Organism
      - Enzyme *isa* Biologically Active Substance
    - among relations
      - treats *isa* affects
  - non-hierarchical
    - Sign or Symptom *diagnoses* Pathologic Function
    - Pharmacologic Substance *treats* Pathologic Function

“Biologic Function” hierarchy (isa)



Associative (non-isa) relationships

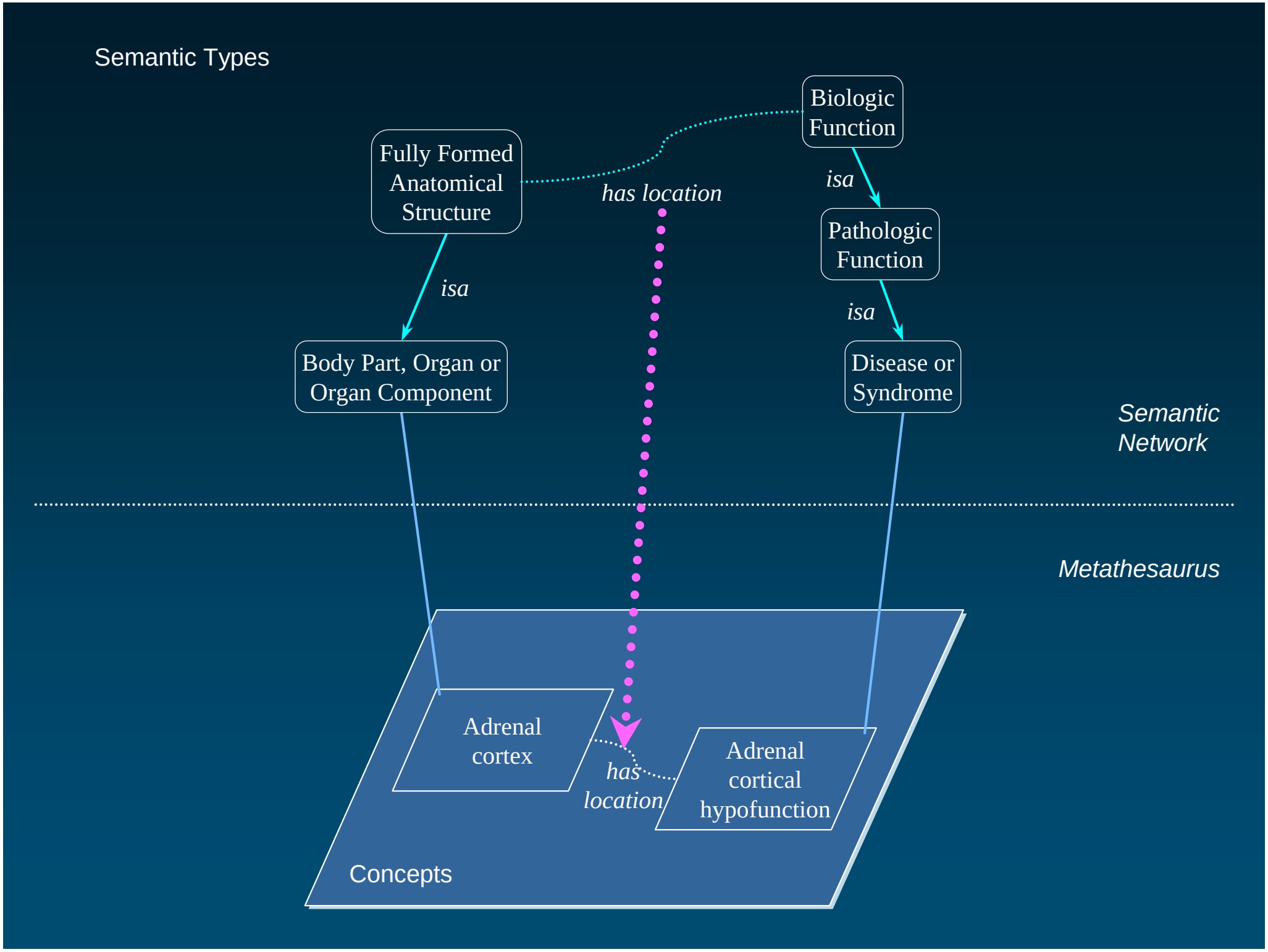


Role

- A relationship between 2 STs is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level
- A child ST inherits properties from its parents (isa relationships)

Applications

- To help qualify inter-concept relationships
  - using the relationships defined between their semantic types in the semantic network
- To strengthen the structure of the Metathesaurus
  - a relationship between 2 concepts should be consistent with the relationships defined between their semantic types in the semantic network
- Semantic interpretation
  - finding semantic relationships between concepts in text


Semantic Types
Biologic Function
Fully Formed Anatomical Structure
has location
isa
Pathologic Function
isa
isa
Body Part, Organ or Organ Component
Disease or Syndrome
Semantic Network
Metathesaurus
Adrenal cortex
has location
Adrenal cortical hypofunction
Concepts

Experiment

[McCray & al. (in press)]

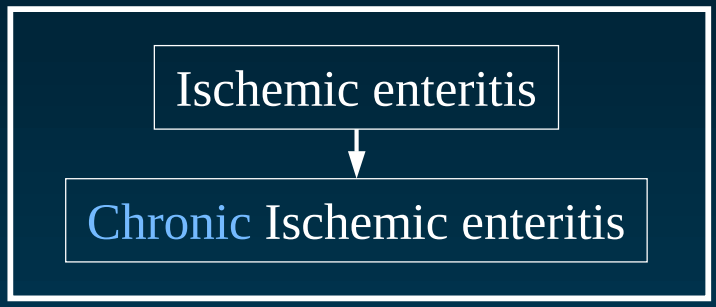
- 3764 concepts related to Heart
- 6894 pairs of related concepts
  - A relation can be inferred unambiguously from the Semantic Network (65%)
  - Multiple semantic links possible (22%)
  - Violation of the Semantic Network (13%)
    - Wrong inter-concept relationship
    - Wrong categorization
    - Both

From concept spaces to ontologies

3. Linguistic approach

Modifiers and relations (1)

- Adjectival modification generally induces hyponymy



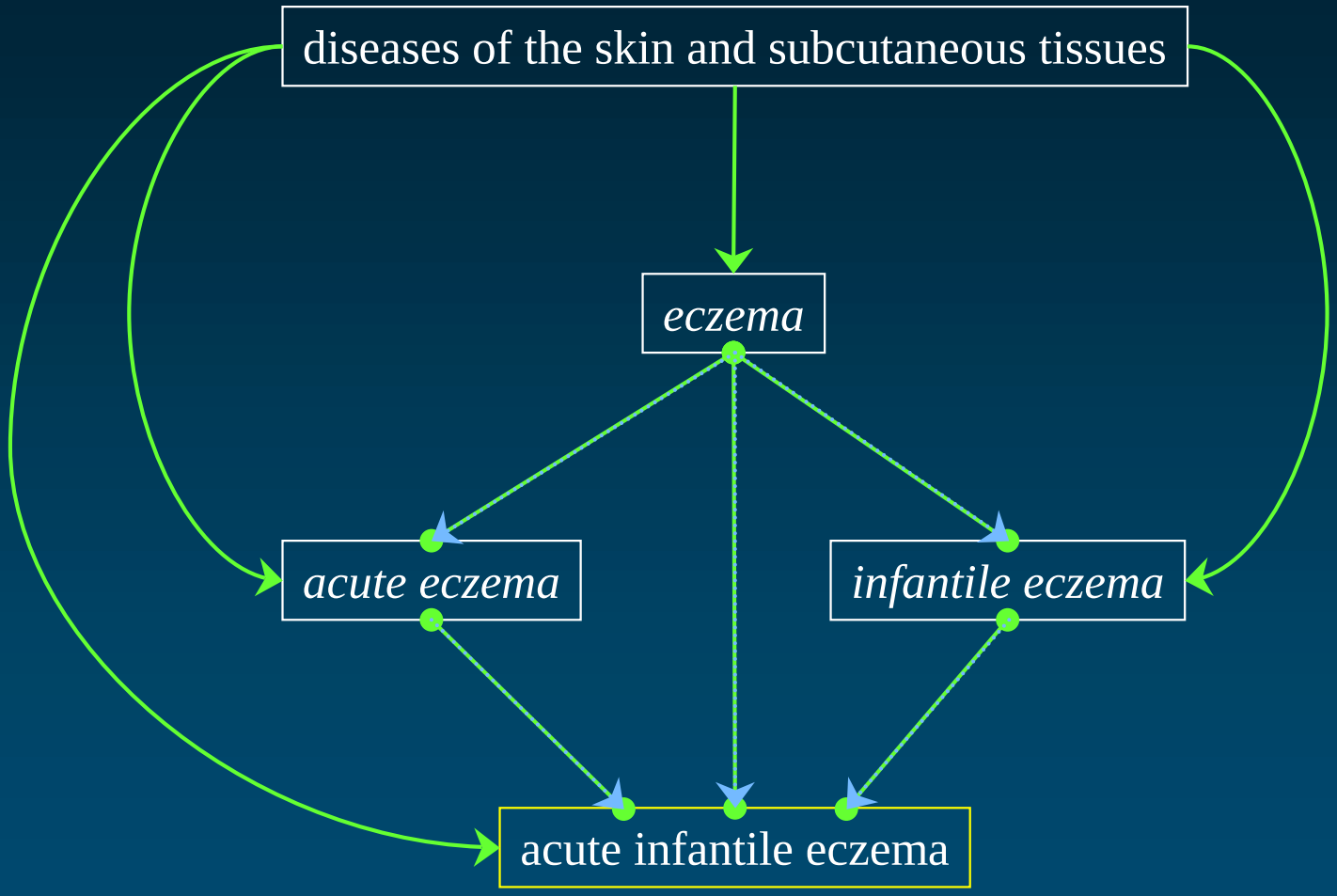
- Names “X” and “mod X”
  - Should name different concepts
  - Concept “X” should be a supertype of “mod X” (assuming that “X” and “mod X” have a common supertype)
- Applications
  - Identify relationships among concepts (in a terminology)

Experiment

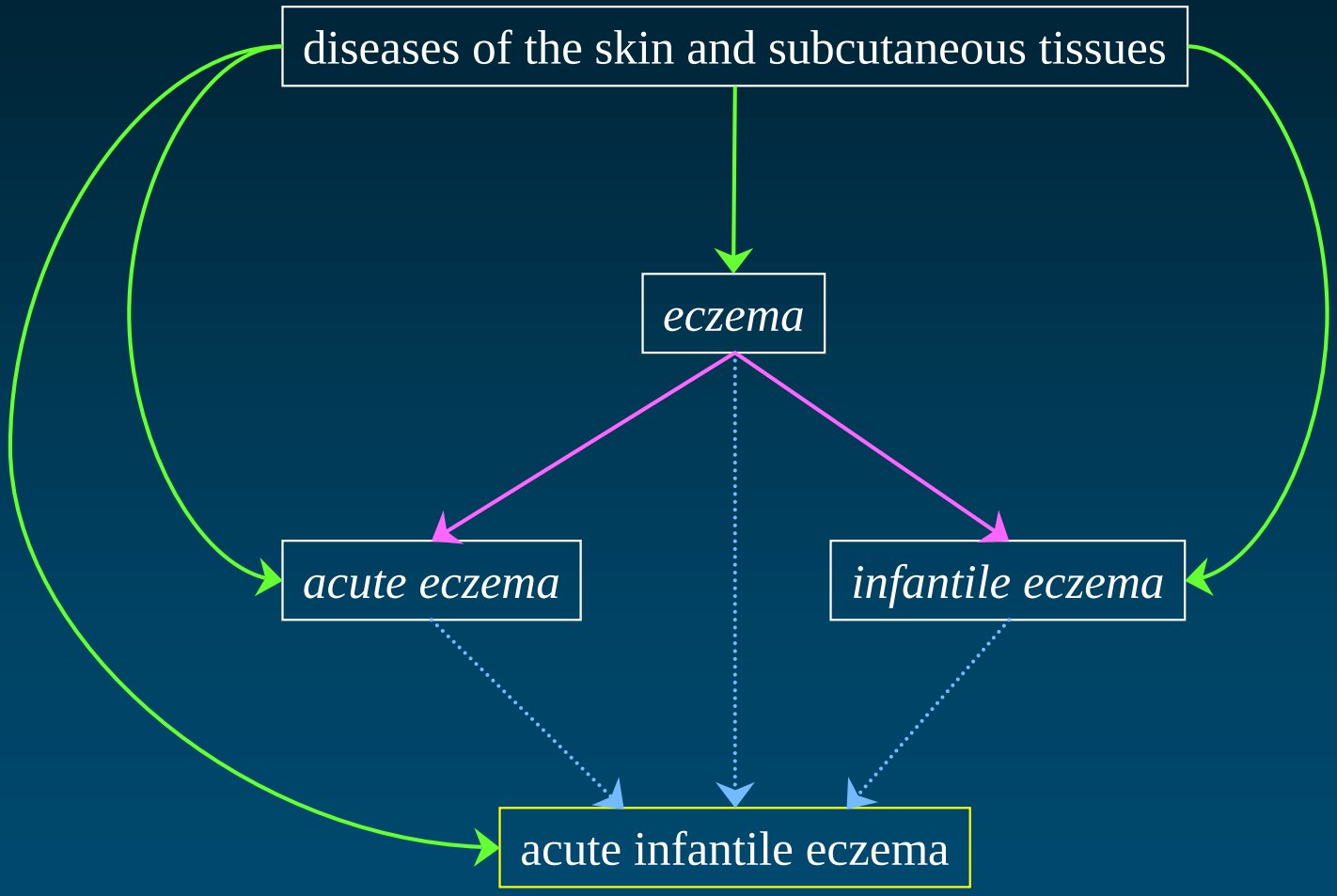
[Bodenreider & al., TIA 2001]

- 28,851 pairs of terms
  - Original SNOMED term
  - Transformed term (found in UMLS)
- Corresponding relationship in the Metathesaurus
  - Hierarchical in 50% of the cases
  - « Sibling » in 25% of the cases
  - Missing in 25% of the cases

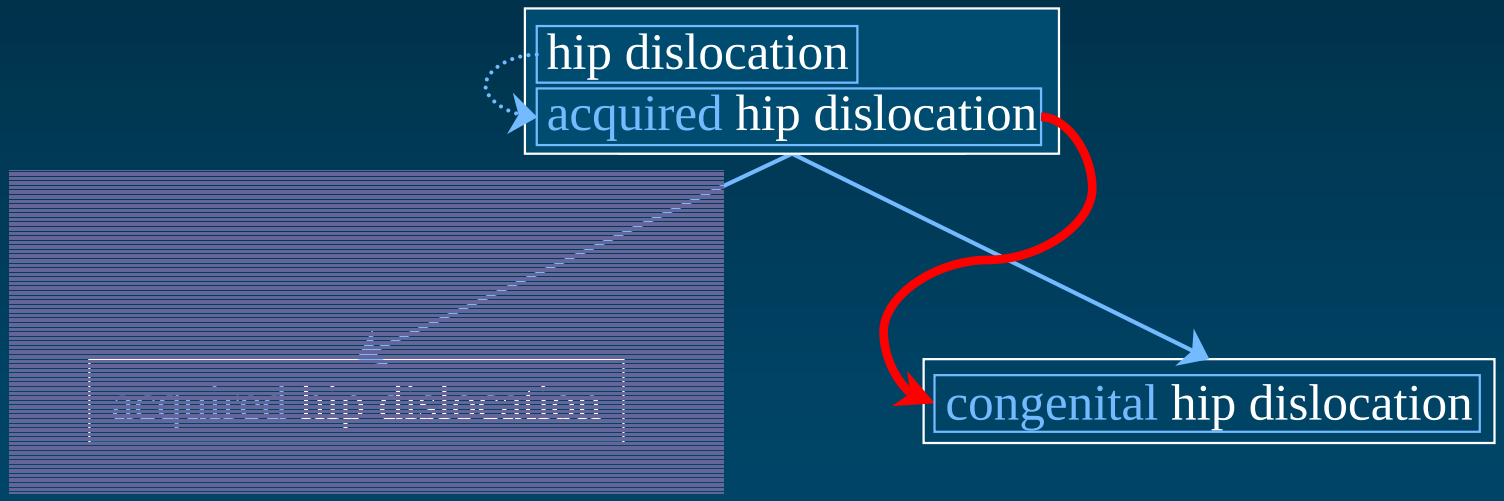
Lack of structure within a source



Lack of links across vocabularies



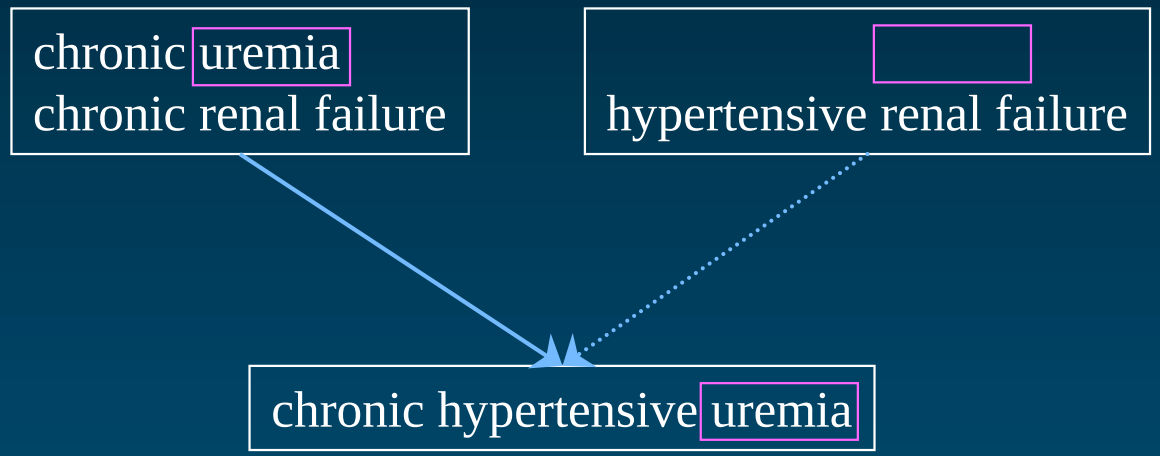
Underspecified terms



Plesionymy

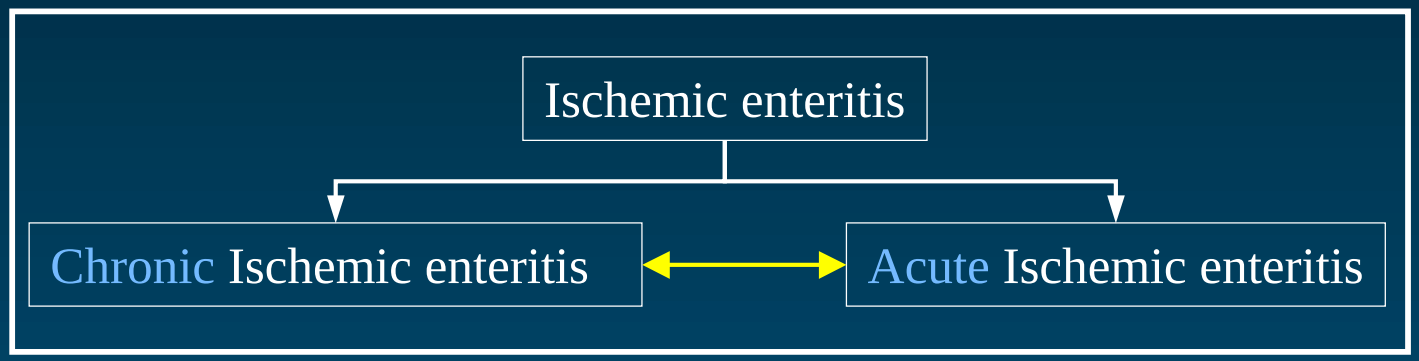

posttransfusion hepatitis
posttransfusion viral hepatitis

Missing synonymy



Modifiers and relations (2)

- Opposite modifiers modifiying a term “X”
  - Should name concepts distinct from “X”
  - Should name subtypes of “X”



- Applications
  - Assess consistency

Experiment

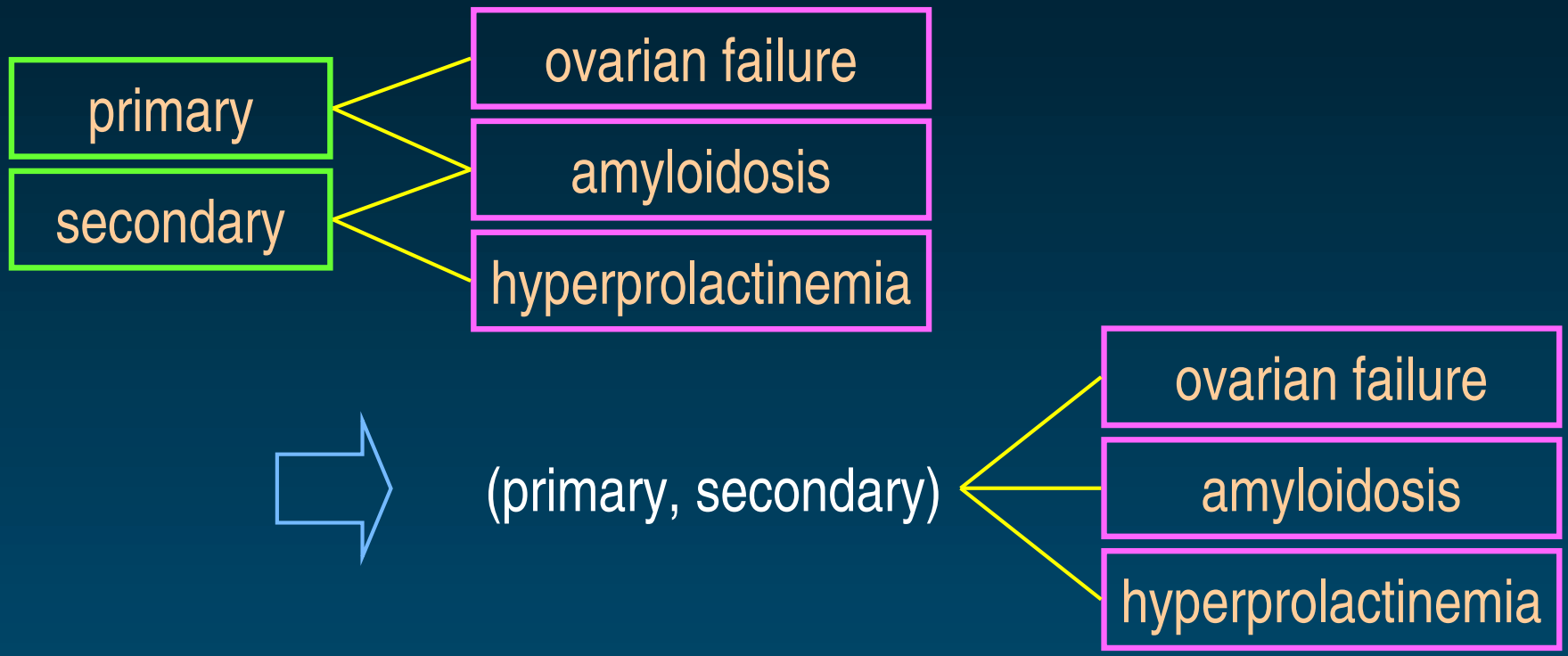
[Bodenreider & al., NLPBA 2002]

- 4 pairs of frequently occurring opposite modifiers

(acute, chronic)
(unilateral, bilateral)
(primary, secondary)
(acquired, congenital)

- Elements studied
  - Presence of the terms and their context
  - Relationships between the 2 modified terms
  - Relationships between each term and its context

Method Transforming terms



primary ovarian failure
primary amyloidosis
primary hyperprolactinemia

secondary ovarian failure
secondary amyloidosis
secondary hyperprolactinemia

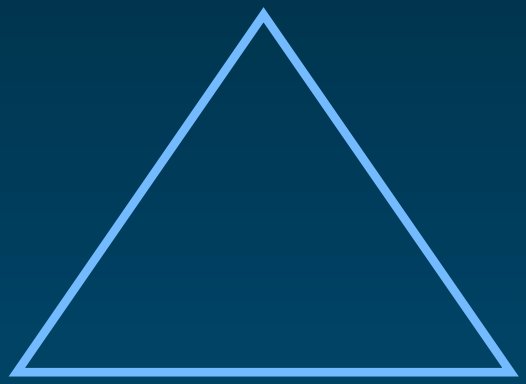
Results Acquired/congenital

		UMLS	
present	Both (e.g.,acquired keratodermia + congenital keratodermia)	97	10%
	« Acquired » term only	76	8%
	« Congenital » term only	801	82%
	Context term (e.g., keratodermia)	418	43%
« Acquired » and « congenital » terms siblings		51	5%
Relationship Of « Acquired » Or « congenital » terms To context	child	181	17%
	siblings	93	9%
	synonyms	82	8%
	none	715	67%

Discussion

Is the concept represented in the ontology?

concept



symbol

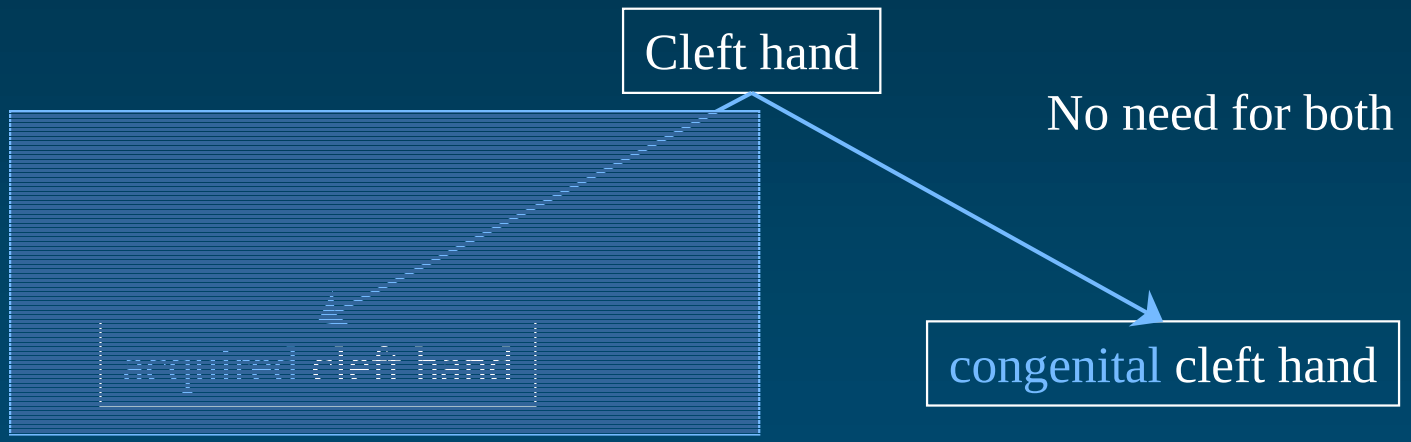
Is the term present in the terminology?

referent

Does the referent exist in the world?

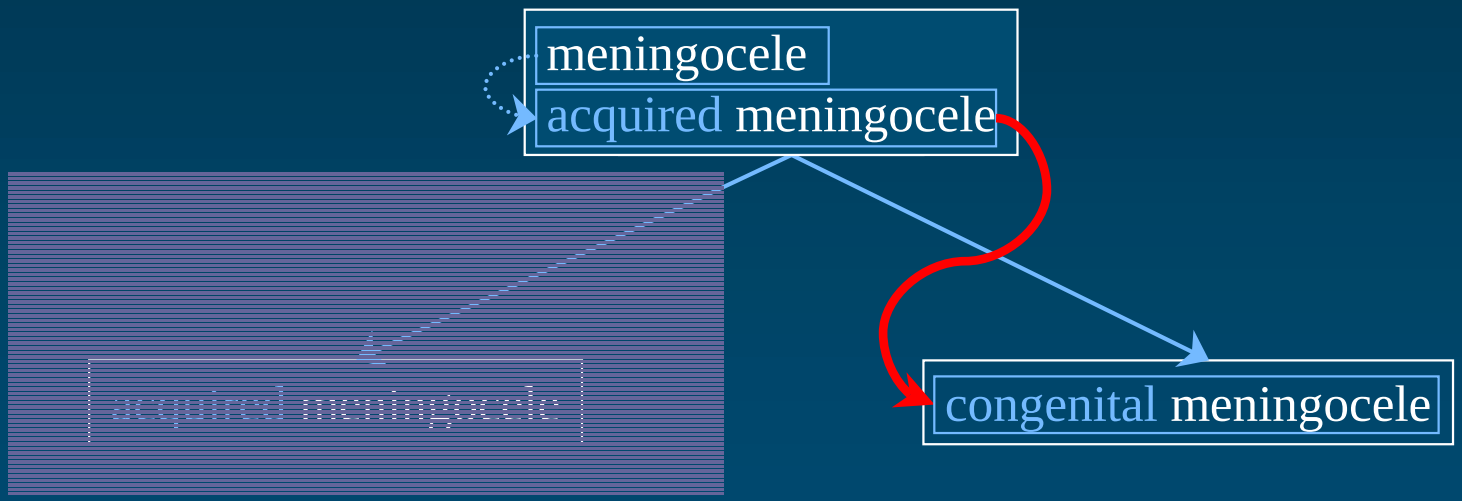
Missing referent

- We artificially created terms by associating modifiers with context
- Medical knowledge



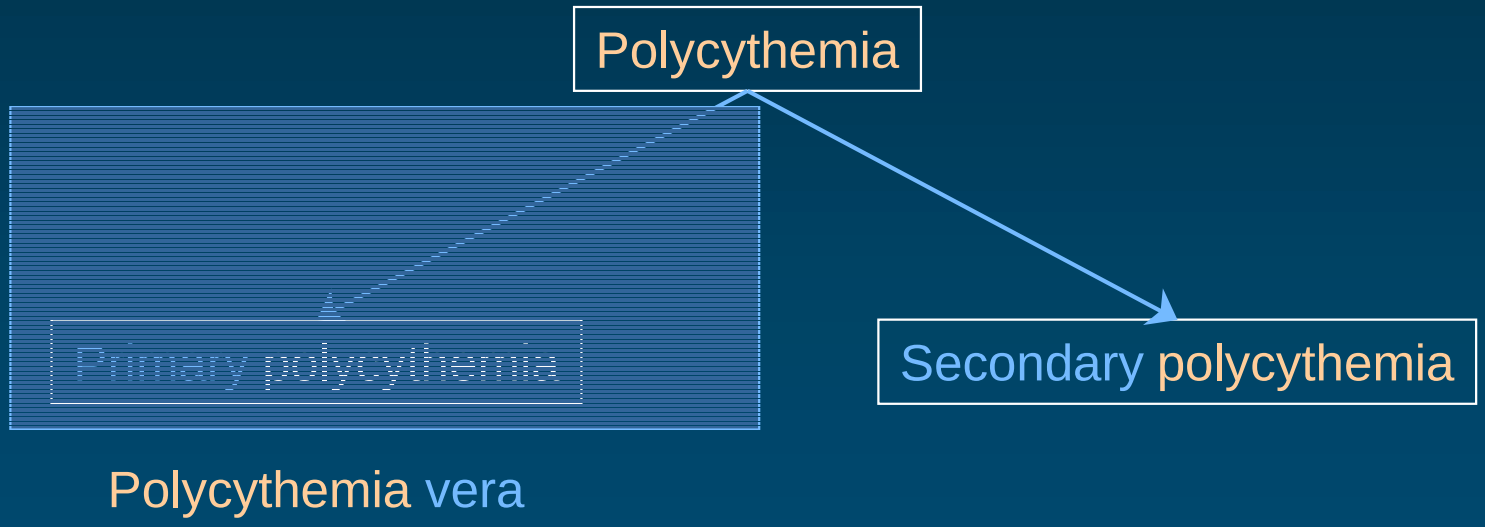
Missing concept

- Knowledge representation, knowledge acquisition
- Distinction among concepts
- Typical form



Missing symbol

- Lexical knowledge
- Synonymy



Modifiers and relations (3)

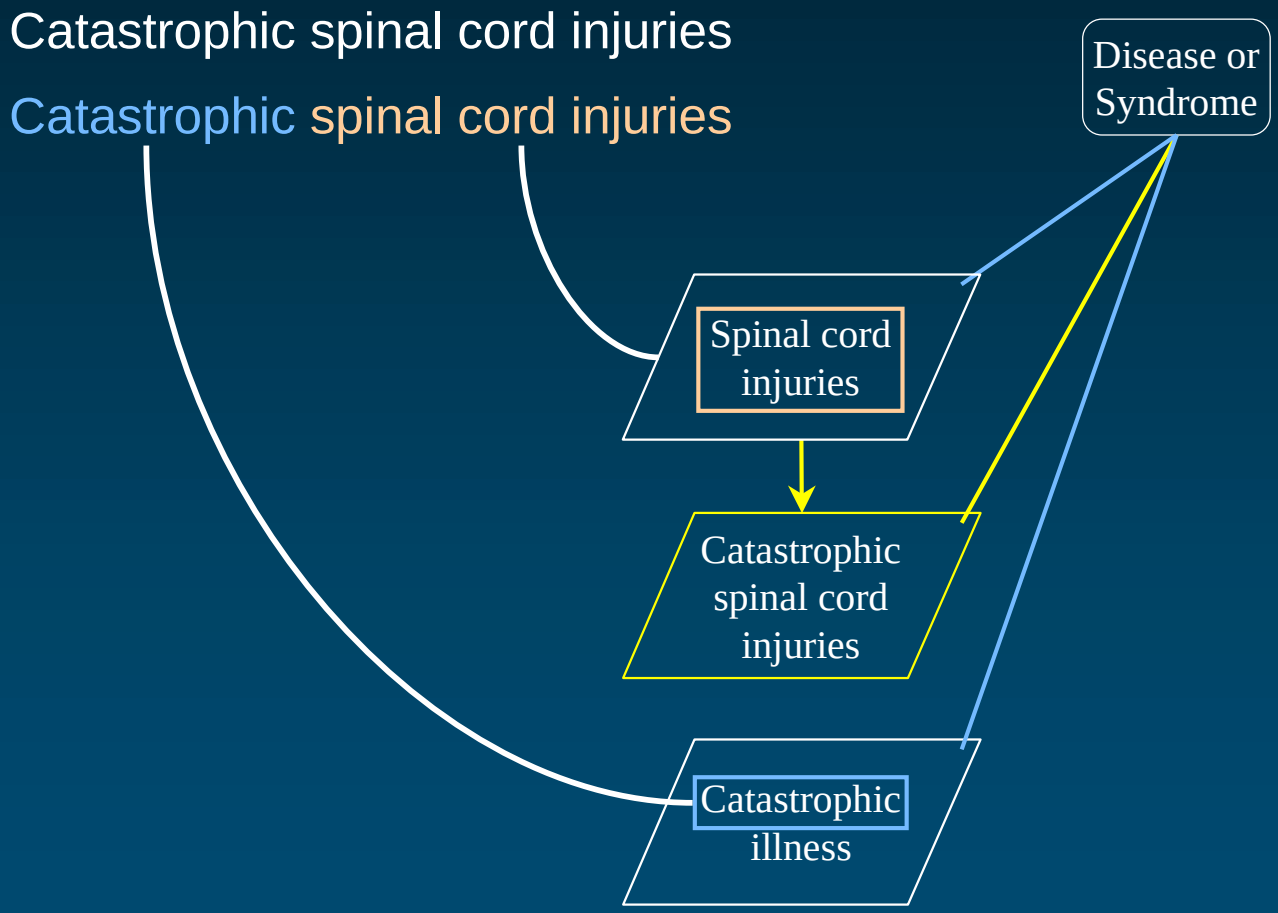
- Modifiers should induce the same kind of relation when applied to different terms “X” and “Y”
  - If rel(“mod X” ,“X”)
  - And “X” and “Y” subtypes of “Z”
  - Then rel(“mod Y”, “Y”)
- Application
  - Extend an existing terminology/ontology with a corpus
    - searching for “mod Y”
    - while knowing “mod X” and “Y”

Experiment

[Bodenreider & al., NAACL 2002]

- 3 million simple noun phrases extracted from MEDLINE citations
- 125,000 new terms identified and associated with an existing UMLS concept
- 83% of the associations are relevant (sample reviewed manually)

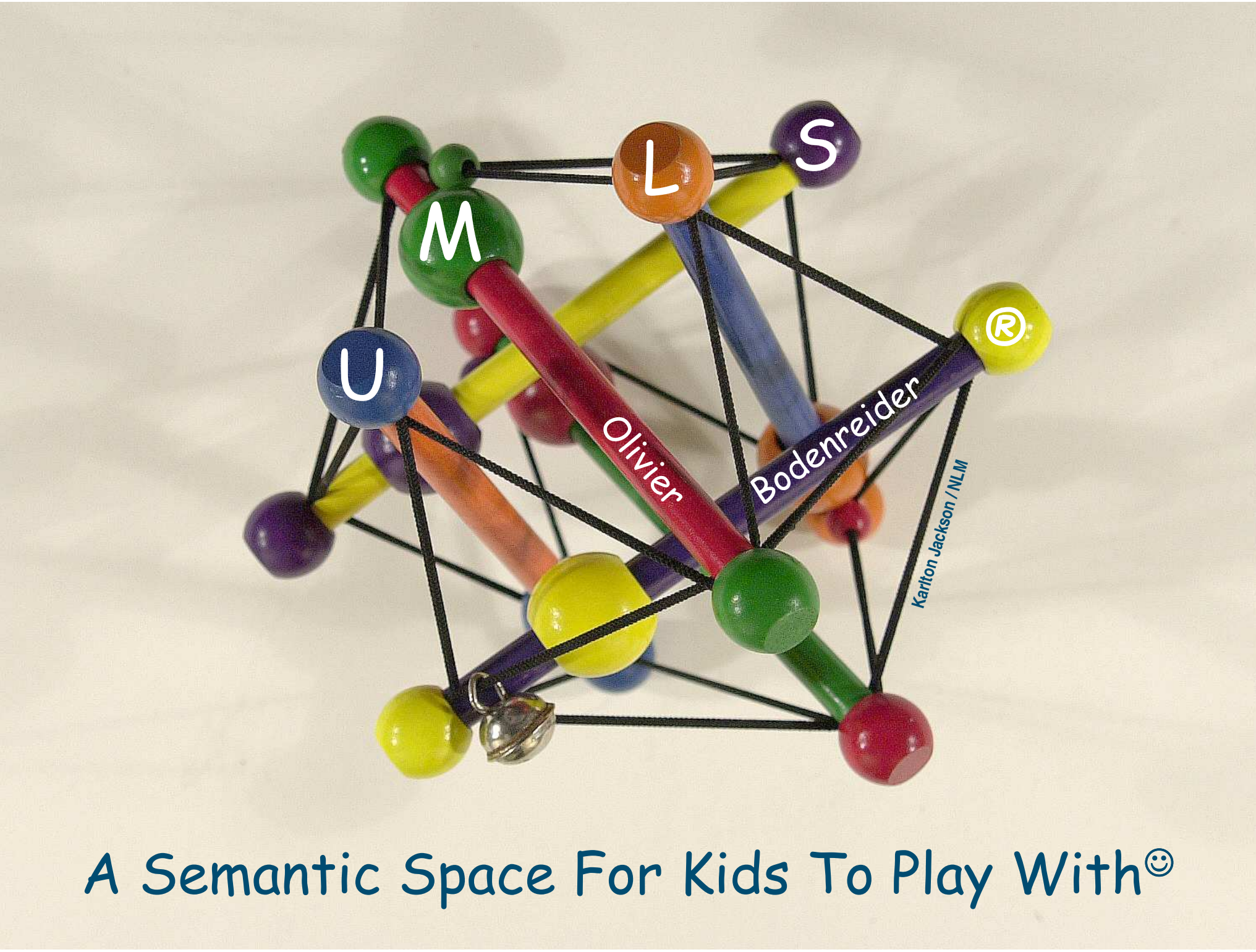
Example


Catastrophic spinal cord injuries
Catastrophic spinal cord injuries
Disease or Syndrome
Spinal cord injuries
Catastrophic spinal cord injuries
Catastrophic illness

Conclusions

Conclusions

- The UMLS certainly has the potential to provide [most of] the concepts and relationships needed in a biomedical ontology
- However, additional effort is needed for selecting relationships meeting ontological requirements
- Comparison to other knowledge sources would also be helpful
  - Medical (GALEN, SNOMED-RT/CT)
  - General (Cyc, WordNet)
  - Specialized (GeneOntology)



A Semantic Space For Kids To Play With©

Contact information

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

olivier@nlm.nih.gov

Anita Burgun

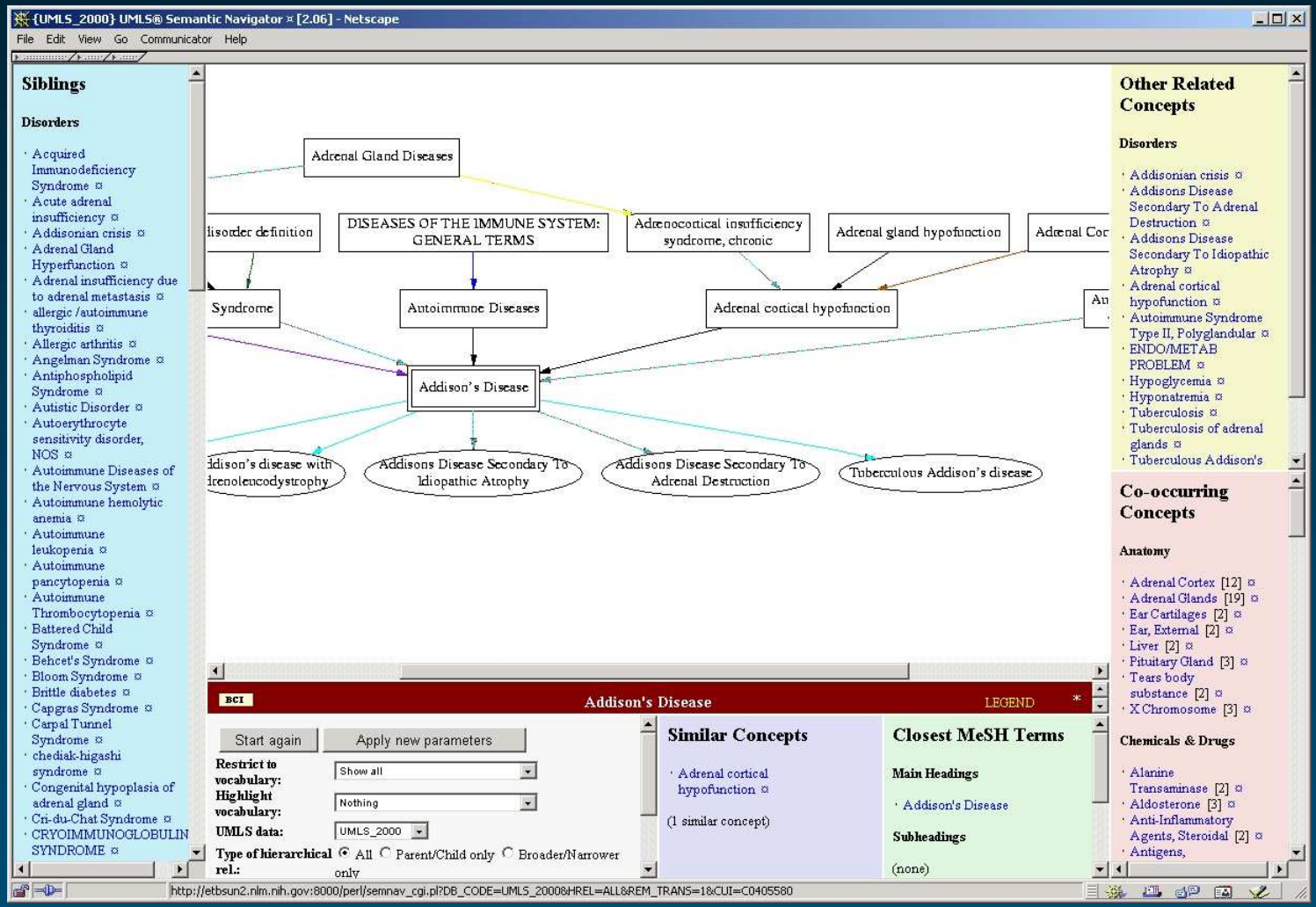
Laboratoire d'Informatique Médicale
Université Rennes 1
Rennes - France

Anita.Burgun@univ-rennes1.fr

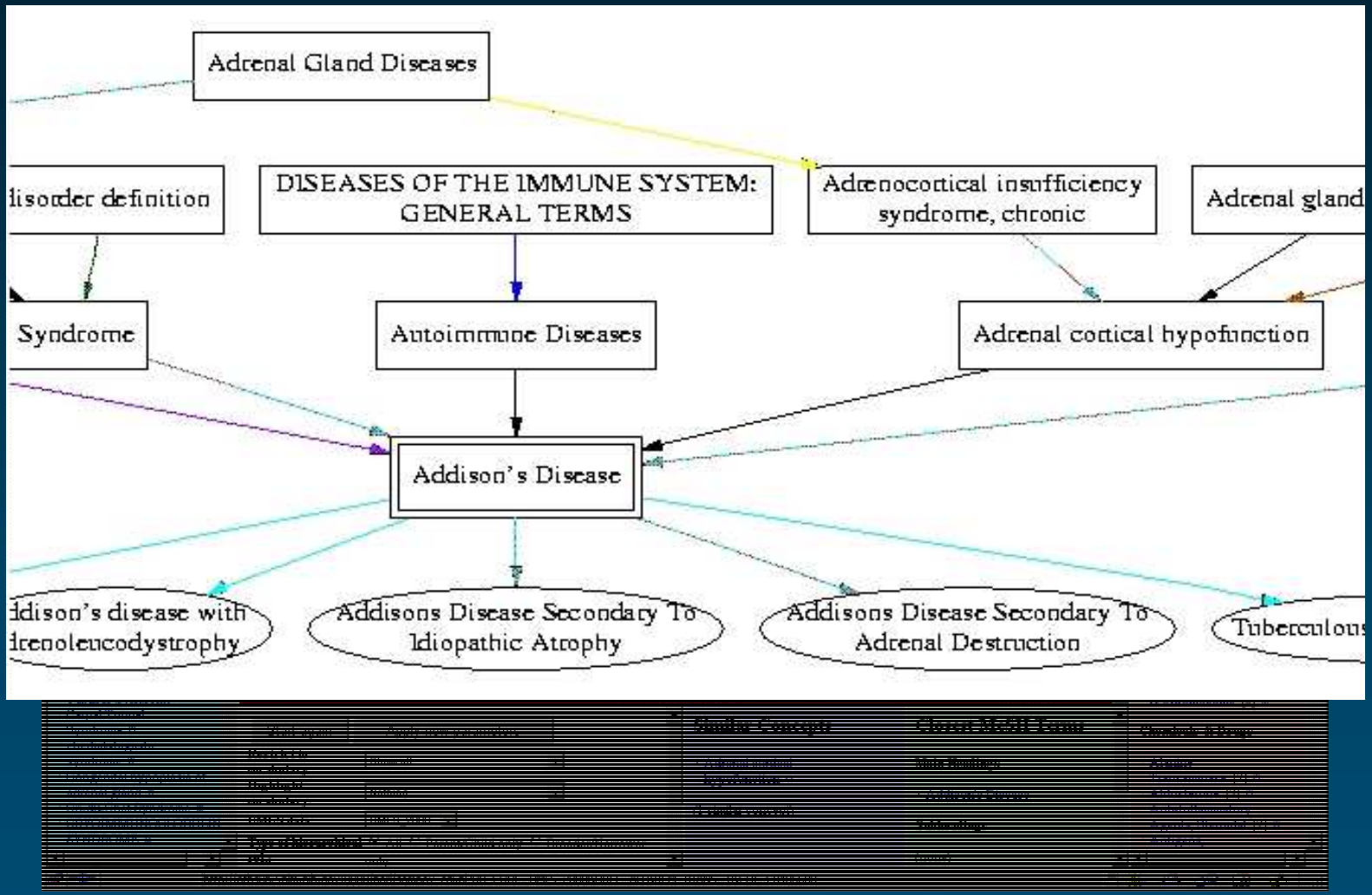
Appendix 1

UMLS Semantic Navigator

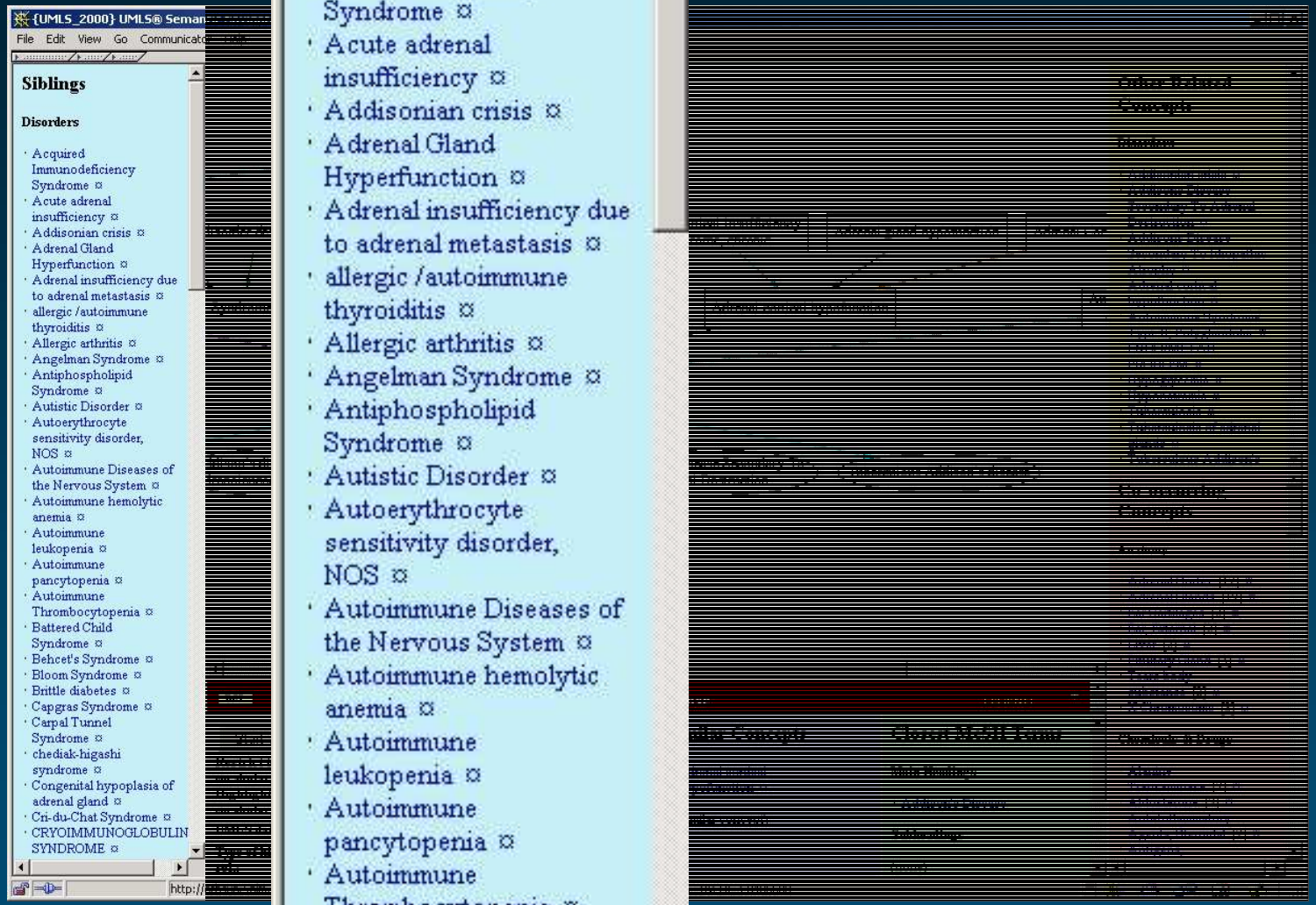
UMLS Semantic Navigator



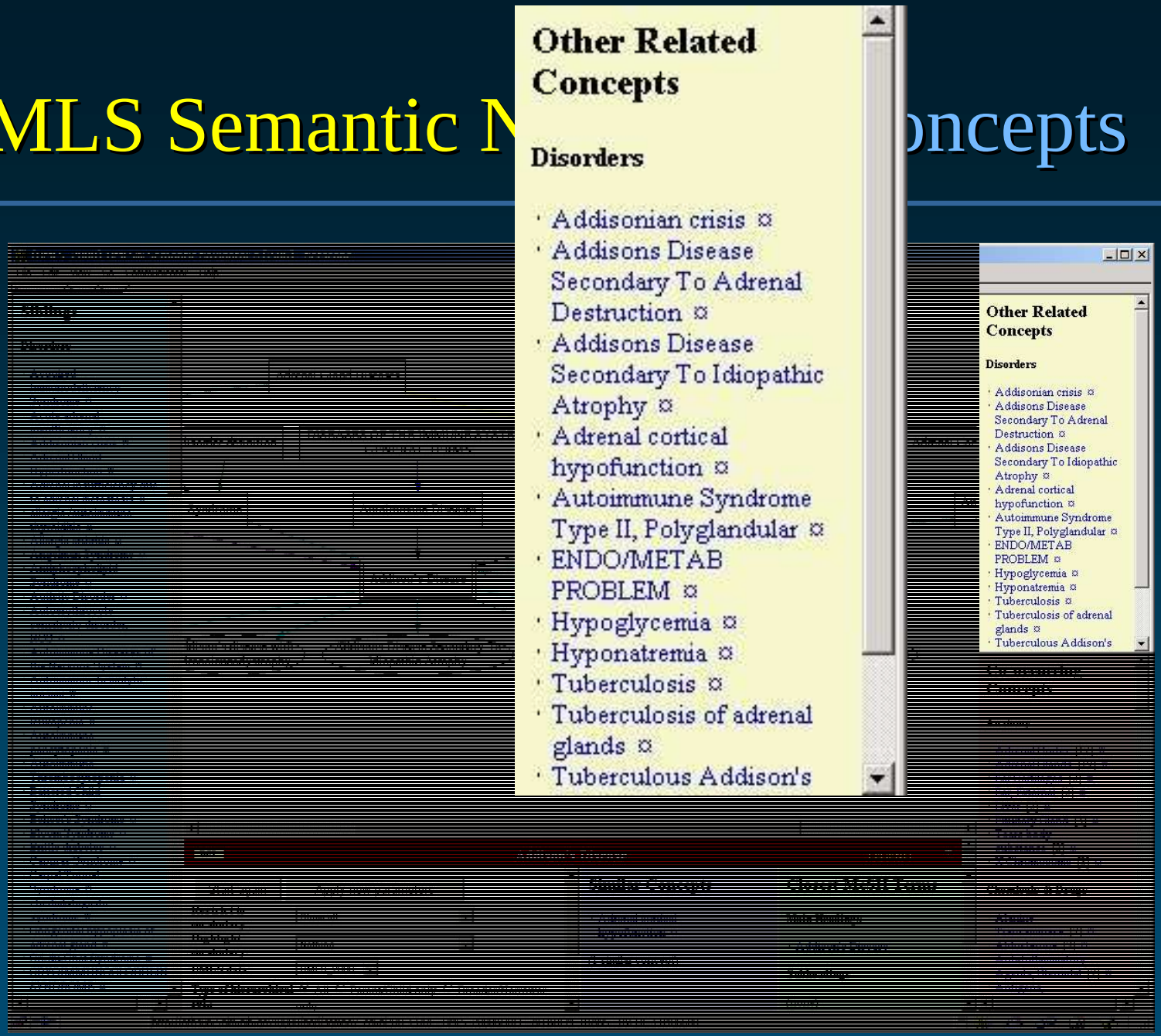
UMLS Semantic Navigator Concepts



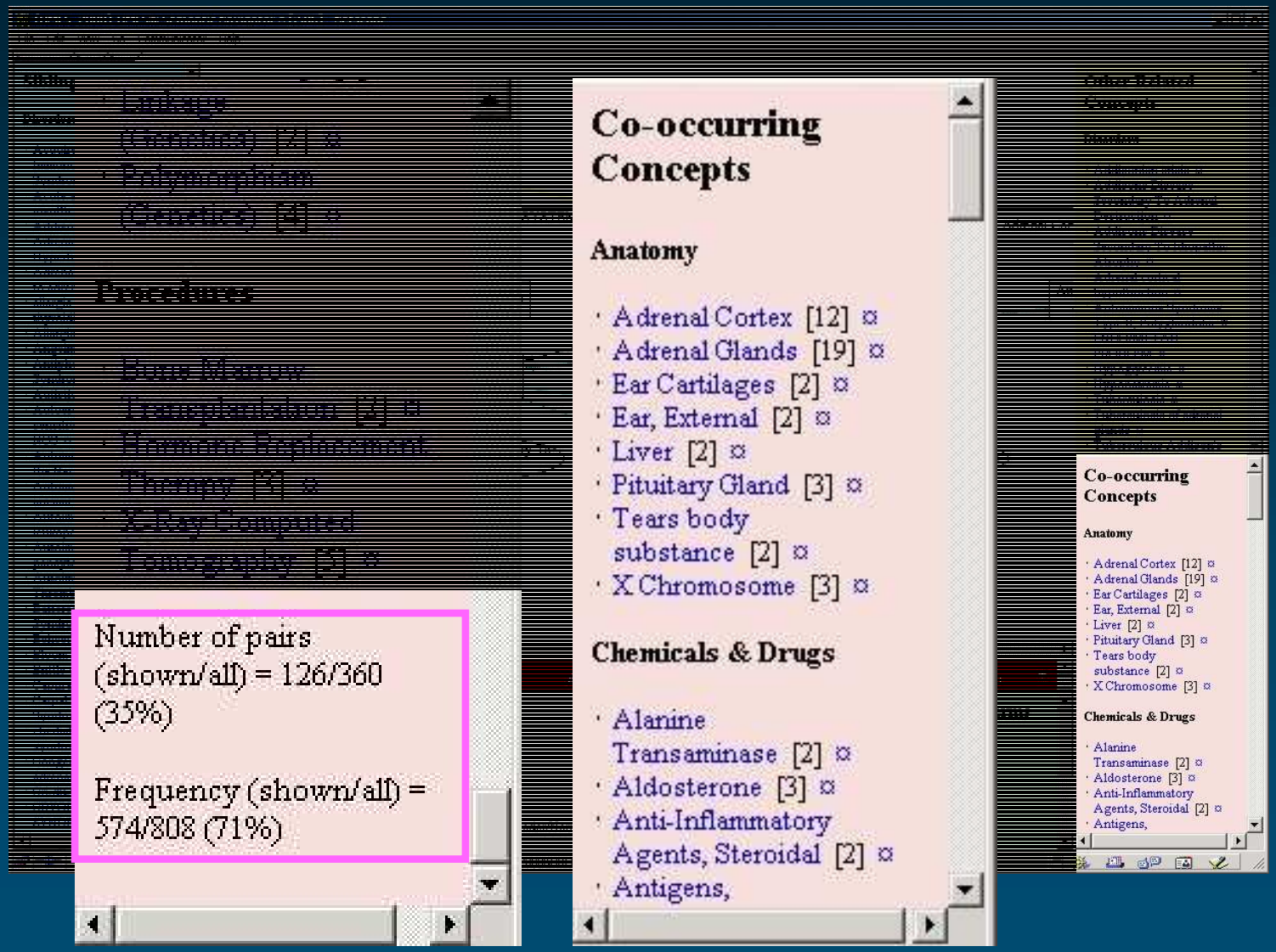
UMLS Se igator Concepts



UMLS Semantic N oncepts



UMLS Semantic Navigator Concepts



UMLS Semantic N